

**Direction des Études et Synthèses Économiques**

**G 2006 / 07**

**Le rôle des conditions de travail  
dans les absences pour maladie**

**Cédric AFSA et Pauline GIVORD**

**Document de travail**



**Institut National de la Statistique et des Études Économiques**

# INSTITUT NATIONAL DE LA STATISTIQUE ET DES ÉTUDES ÉCONOMIQUES

*Série des documents de travail  
de la Direction des Etudes et Synthèses Économiques*

**G 2006 / 07**

## **Le rôle des conditions de travail dans les absences pour maladie**

**Cédric AFSA\* et Pauline GIVORD\*\***

SEPTEMBRE 2006

Les auteurs remercient les participants au séminaire interne du CREST, au séminaire du Département des Etudes Economiques d'Ensemble (et tout particulièrement Eve Caroli, PSE, qui a accepté d'y discuter une version préliminaire de l'étude), aux XIIIèmes Journées du Longitudinal d'Aix-en-Provence, et aux XXIIIèmes Journées de Micro-économie Appliquée de Nantes (où Ali Skalli, ERMES, a assuré la discussion). Ils restent seuls responsables des erreurs qui subsisteraient.

---

\* Département des Etudes Economiques d'Ensemble - Division « Redistribution et Politiques Sociales »  
Timbre G210 - 15, bd Gabriel Péri - BP 100 - 92244 MALAKOFF CEDEX

\*\* CREST - Département de la Recherche 15, bd Gabriel Péri - 92245 MALAKOFF Cedex

## Le rôle des conditions de travail dans les absences pour maladie

### Résumé

Les économistes se sont rarement intéressés au rôle des conditions de travail dans les absences pour maladie. Pourtant, de mauvaises conditions de travail sont susceptibles d'influencer les comportements d'activité que cet effet passe par l'état de santé ou par la propension à l'absentéisme à état de santé donné.

Pour préciser ces liens, nous commençons par l'examen d'un modèle théorique qui prend explicitement en compte l'état de santé du salarié et son évolution, afin de démêler les liens entre conditions de travail, santé et comportement d'activité. Ce modèle permet de mettre en avant deux effets contradictoires : de mauvaises conditions de travail dégradent l'état de santé et accroissent les absences pour maladie, mais l'absentéisme peut être inversement freiné par un effet salaire, si les conditions de travail défavorables sont compensées par un salaire plus élevé. Dans ces conditions, déterminer lequel de ces deux effets l'emporte devient un problème empirique.

Nous nous intéressons spécifiquement à l'impact de l'irrégularité des horaires sur les arrêts maladie pour les ouvriers de sexe masculin travaillant dans le secteur privé. Pour l'estimer, nous utilisons une méthode d'appariement sur le score de propension. Nous contrôlons la robustesse de ses résultats en utilisant une méthode alternative de « sélection sur inobservables ». Les principales conclusions sont les suivantes. L'irrégularité des horaires joue un rôle significatif dans les absences pour maladie : elle serait responsable d'un cinquième des arrêts maladie pris par les ouvriers en horaires irréguliers. Ce pourcentage est nettement plus élevé que celui obtenu en comparant directement les arrêts maladie des ouvriers travaillant en horaires irréguliers et de ceux en horaires réguliers. Néanmoins, le sens et l'importance de l'impact de l'irrégularité des horaires varient avec l'âge du salarié.

**Mots-clés** : conditions de travail ; santé ; arrêts maladie ; horaires de travail ; méthode d'appariement.

## The role of working conditions in sickness absence

### Abstract

Economists have paid very little attention to the role of working conditions in sickness absence. Yet, bad working conditions are a potential determinant of labour supply, either directly or through their impact on health. This study tries to shed some light on this issue.

To begin with, we examine a model of labour supply which explicitly takes into account the employee's health status and its evolution. This model provides a useful framework in order to interpret and disentangle the links between working conditions, health status and work. In particular it shows that two effects occur: bad working conditions have a negative impact on health and labour supply. But this can be compensated by a wage effect, if bad working conditions are compensated by higher wages. Thus the impact of working conditions on absenteeism is theoretically ambiguous. This requires an empirical estimation in order to determine which effect prevails.

We specifically study the impact of working irregular schedules on absenteeism for male manual workers in private sector. We use a propensity score matching method. In order to check the robustness of our estimates we also use a "selection on unobservables" specification. Our estimates show that working irregular schedules has a significant impact on sickness absence: roughly 20 % of the employees with irregular schedules who were absent for sickness reason would not have stopped working if their schedules were regular. This effect is definitely more important than that estimated by simple descriptive statistics. Nevertheless, the sign and the extend of the effect depend on age.

**Keywords** : working conditions ; health demand ; absenteeism ; work schedules ; matching.

**JEL Classification** : J28; I12 ; J22 ; J81.

## Sommaire

<b>Introduction .....</b>	<b>5</b>
<b>I - Santé, travail et absence : une revue de littérature .....</b>	<b>7</b>
<b>II - Le cadre théorique.....</b>	<b>11</b>
<i>II.1 Le modèle de base</i>	<i>11</i>
<i>II.2 L'introduction des conditions de travail dans le modèle</i>	<i>13</i>
<b>III - Santé et conditions de travail : les problèmes empiriques.....</b>	<b>15</b>
<i>III.1 Recueillir des données sur les conditions de travail             et la santé : un processus de collecte exigeant</i>	<i>15</i>
<i>III.2 La source retenue</i>	<i>16</i>
<i>III.3 Quelques statistiques sur l'irrégularité des horaires             et l'absence pour maladie</i>	<i>18</i>
<b>IV - Méthode économétrique .....</b>	<b>21</b>
<i>IV.1 Méthode d'appariement</i>	<i>21</i>
IV.1.1 Principe et hypothèses de la méthode	21
IV.1.2 Le score de propension	22
IV.1.3 L'estimation du contrefactuel	23
IV.1.4 Le choix des variables explicatives	25
IV.1.5 En guise de résumé	25
<i>IV.2 Une estimation alternative de l'impact des horaires irréguliers</i>	<i>26</i>
<b>V - Résultats .....</b>	<b>29</b>
<i>V.1 Résultats sur l'ensemble des ouvriers en horaires irréguliers             (tous âges confondus)</i>	<i>29</i>
<i>V.2 Résultats par âge</i>	<i>31</i>
<b>VI - Quelques remarques conclusives .....</b>	<b>33</b>
<b>Références .....</b>	<b>35</b>
<b>Annexe : les macros SAS.....</b>	<b>38</b>



## Introduction

Les économistes se sont rarement intéressés à l'influence des conditions de travail sur les comportements d'activité des salariés, alors que cette question fait l'objet de nombreux travaux dans d'autres disciplines des sciences sociales. Elle est pourtant au cœur de préoccupations de politique publique. C'est, par exemple, un élément du débat sur le faible taux d'emploi des seniors. Ainsi, dans leur rapport pour le Conseil d'Analyse Economique, d'Autume *et alii* (2005) rappellent le rôle joué par la pénibilité du travail dans les comportements d'activité, celui d'user prématurément les salariés qui subissent les plus fortes contraintes et de les inciter à des départs précoces. Ils préconisent d'améliorer les conditions de travail des seniors en aménageant leurs tâches et leurs horaires. Des considérations de cette nature sont reprises au niveau européen. Par exemple, une communication de la Commission Européenne s'est appuyée sur un des objectifs du Conseil de Lisbonne de mars 2000 - celui d'augmenter l'emploi en quantité et en qualité - pour faire de la qualité du travail une dimension importante du développement économique et social (Commission Européenne, 2002). Une autre préoccupation - récurrente - des pouvoirs publics en France est la croissance des indemnités journalières de maladie. Dans l'état des lieux que la Caisse Nationale d'Assurance Maladie a dressé en 2005 sur la question, les conditions de travail faisaient partie de la liste des déterminants des arrêts maladie (Midy, 2005). Mais le diagnostic reste partiel et demande une analyse plus approfondie.

C'est précisément le lien entre conditions de travail et absence pour maladie que nous cherchons ici à établir. Pour ce faire, nous commençons par construire un modèle théorique selon lequel la pénibilité du travail influe sur le comportement d'absence par son effet sur l'état de santé du salarié. Le modèle montre que si la pénibilité du travail donne lieu à une compensation salariale même partielle, alors son impact sur les arrêts maladie est a priori ambigu. Il est la résultante de deux effets antagonistes : l'effet incitatif au travail d'une meilleure rémunération est contrebalancé par l'effet négatif de la détérioration de l'état de santé. Dans ces conditions, la question du lien entre conditions de travail et absences pour maladie devient une question empirique.

Pour la traiter, nous nous intéressons à un aspect particulier des conditions de travail, qui est le rythme du travail. Plus exactement, nous isolons les salariés travaillant en horaires alternés ou variables d'une semaine à l'autre, et les opposons à ceux dont les horaires sont les mêmes chaque semaine. Cette caractéristique du poste de travail a l'avantage, contrairement à d'autres aspects des conditions de travail, d'être assez facilement objectivable. Nous cherchons à évaluer l'impact que le seul fait de travailler avec ces horaires « irréguliers » a sur la probabilité de s'arrêter pour maladie. Leur impact sur la santé a été documenté par des travaux en ergonomie ou en épidémiologie, mais il n'existe pas à notre connaissance de conclusions tranchées sur le lien entre irrégularité des horaires et absence pour maladie. Nous mobilisons les données de l'enquête sur l'Emploi de l'Insee, en nous centrant sur une population relativement homogène, celle des ouvriers hommes du secteur privé. A partir du moment où les ouvriers travaillant en horaires irréguliers se distinguent des autres, il nous faut traiter les biais potentiels liés à ces éventuels effets de sélection dans les postes de travail. Nous le faisons, du moins en partie, en utilisant une méthode d'appariement par le score de propension. Nos estimations montrent que : (a) travailler avec des horaires irréguliers a un impact très significatif sur la probabilité de s'arrêter pour maladie ; (b) cet effet est nettement plus important que celui estimé par de simples statistiques descriptives ; et (c) le sens de l'effet et son amplitude varient beaucoup avec l'âge.



## I - Santé, travail et absence : une revue de littérature

Notre étude se situe à l'intersection de trois domaines : les conditions de travail, la santé et l'absentéisme. A notre connaissance, les tentatives d'intégrer ces trois dimensions dans une approche économique sont pour l'instant très rares.

Comme le soulignent Brown et Sessions (1996) dans leur synthèse, la littérature économique est longtemps restée discrète sur l'étude de l'absentéisme et de ses causes, contrairement à d'autres disciplines des sciences sociales. Un premier courant, datant des années 1970, s'est appuyé sur le modèle standard d'arbitrage travail/loisir et fait dépendre la propension à être absent de l'écart entre, d'une part, la durée contractuelle liant le salarié à son employeur et, d'autre part, la durée optimale pour l'individu, celle maximisant son utilité sous contrainte budgétaire. Une prédiction de ce modèle est qu'une augmentation des heures contractuelles accroît la propension à s'absenter. Les travaux empiriques sur cette idée théorique se sont développés au cours des années 1980 (voir par exemple Dunn et Youngblood, 1986).

Un second courant a pris comme point de départ le modèle dit du « tire-au-flanc » (*shirking model*) développé par Shapiro et Stiglitz (1984). Le salarié choisit le niveau d'effort qu'il doit fournir pour accomplir ses tâches. Le « tire-au-flanc » est celui qui maximise son utilité en fournissant l'effort minimal. Les travaux sur l'absentéisme qui se fondent sur ce modèle ont en commun de considérer l'absence comme révélant le niveau d'effort fourni par le salarié. Dans ce contexte, et à partir du moment où l'employeur ne peut pas le connaître avec certitude, le problème peut se poser en termes d'aléa moral. Plusieurs auteurs l'ont fait et ont alors centré leurs analyses sur les moyens d'en limiter les effets supposés. La piste généralement privilégiée a été le durcissement des règles de droit aux arrêts maladie, et plus précisément la diminution du taux de compensation salariale (c'est-à-dire le montant des indemnités maladie rapporté au salaire). Une série de travaux empiriques en ont confirmé l'effet attendu, certains utilisant des « expériences naturelles » (par exemple, Meyer *et alii*, 1995, Bolduc *et alii*, 2002, ou Andren, 2005).

Cette approche par l'aléa moral combinée à des modèles de négociation a fourni le cadre théorique permettant d'expliquer les phénomènes cycliques des arrêts maladie par la variation du pouvoir de négociation des salariés vis-à-vis de leur employeur : en phase basse de l'activité économique, en présence d'un chômage élevé, les salariés hésiteraient davantage à déclarer des arrêts maladie par crainte d'être licenciés. Dans le même ordre d'idées, Engelland et Riphahn (2005) ont montré que les salariés sous contrat à durée indéterminée étaient plus souvent absents que ceux sous contrat temporaire.

Dans leur étude comparative sur l'absentéisme dans 12 pays européens, Frick et Malo (2005) ont précisément introduit les trois déterminants dont il vient d'être question - le taux d'indemnisation des arrêts maladie, le chômage, le statut du contrat de travail. Ils ont calculé des indicateurs nationaux caractérisant, pour chaque pays, le degré de générosité du système d'indemnisation de l'absence pour maladie, le niveau du chômage et le degré de protection de l'emploi. A côté de ces variables d'environnement institutionnel, ils ont ajouté dans leurs modèles empiriques des caractéristiques proprement individuelles. Les effets des variables d'environnement sont conformes à ce qui est généralement trouvé dans la littérature, mais ils semblent être moins importants que ceux produits par les variables individuelles, comme par exemple l'existence de problèmes de santé au travail.

Ce tout dernier point illustre d'ailleurs un des reproches formulés par Brown et Sessions (1996) à l'encontre des modèles théoriques, surtout lorsque ceux-ci s'intéressent explicitement aux absences pour maladie, qui est d'ignorer l'état de santé. Tout se passe comme si l'absence au travail ne découlait jamais de l'incapacité



de l'individu à travailler mais exprimait seulement son choix de ne pas travailler. Pourtant, en s'en tenant à la seule littérature économique, Allen (1981), par exemple, avait introduit dans son modèle empirique une variable d'auto-évaluation de l'état de santé et en avait montré l'impact significatif sur les absences. Pour répondre à cette critique, Barmby *et alii* (1994) ont enrichi le cadre théorique découlant du modèle de Shapiro et Stiglitz en faisant dépendre la fonction d'utilité du revenu, du loisir et de l'état de santé de l'individu, ce dernier intervenant en réalité comme facteur d'hétérogénéité des préférences..

Cela étant, on ne peut pas faire dépendre les comportements d'absence pour maladie du seul état de santé. On ne peut pas écarter l'hypothèse que ces comportements puissent varier d'un individu à l'autre selon leur degré de motivation, selon leur « goût » de l'effort. C'est d'ailleurs la position de Ose (2005), qui a tenté d'identifier séparément les absences volontaires et celles, involontaires, résultant de problèmes de santé liés à de mauvaises conditions de travail. Toujours dans cet esprit, Grignon et Renaud (2004) ont approfondi le concept de hasard moral en en distinguant deux types : le « pur » hasard moral *ex post* qui est essentiellement sous le contrôle de l'individu, et la forme *ex ante* du hasard moral qui relève davantage de la responsabilité de l'employeur et qui est la variable d'intérêt lorsqu'on étudie l'impact des conditions de travail sur les comportements.

Concernant justement les conditions de travail, elles ont reçu, dans la littérature économique, encore moins d'échos que l'absence. Il faut dire que leur sort a été rapidement « réglé » par la théorie des différences compensatrices, selon laquelle des conditions de travail pénibles sont compensées par des suppléments de salaire (Rosen, 1974). Ceci contraste fortement avec la littérature d'épidémiologie ou d'ergonomie appliquée, où le lien entre conditions de travail et état de santé ou absence pour maladie est abondamment documenté<sup>1</sup> (pour une revue française de littérature récente sur ces questions, voir Lasfargues, 2005).

L'existence de compensations salariales, pour autant qu'elle soit avérée (sur les difficultés empiriques à l'établir, voir Brown, 1980), n'empêche toutefois pas les individus d'être sensibles aux caractéristiques non monétaires de leur emploi. Notamment, des conditions de travail défavorables ou vécues comme telles encouragent la mobilité volontaire (Bonhomme et Jolivet, 2005). Par conséquent, le fait qu'elles puissent jouer sur les comportements sur le marché du travail justifie leur prise en compte dans une approche économique.

Quoi qu'il en soit, les travaux d'économistes centrés sur le rôle des conditions de travail dans les comportements restent rares. Leigh (1991) est une exception<sup>2</sup>. Son travail empirique a conclu à une forte corrélation entre la dangerosité des conditions de travail (mesurée par des indicateurs subjectifs de risque d'accident) et l'absentéisme. Dans le cas français, Askenazy et Caroli (2003) ont étudié les pratiques organisationnelles innovantes, comme la polyvalence ou la flexibilité du temps de travail, et ont montré empiriquement leur impact négatif sur le bien-être des salariés concernés, mesuré par divers indicateurs comme la charge mentale ou encore les accidents du travail.

A notre connaissance, Case et Deaton (2005) sont parmi les premiers à tenter de prendre en compte simultanément les dimensions travail et santé dans un modèle théorique. Pour ce faire, ils ont adapté le modèle de Grossman (1972) sur l'évolution de l'état de santé, qui veut que l'individu investit régulièrement dans son capital santé pour en ralentir la détérioration, mesurée par un coefficient  $\delta$ . Selon leur modèle, la

---

<sup>1</sup> Comme en témoignent les nombreux articles traitant de ces questions dans les revues *Social Science and Medicine* ou *Applied Ergonomics*.

<sup>2</sup> Notons, toutefois, que Leigh a publié son étude dans une revue de médecine appliquée, sous la double compétence d'économiste et d'épidémiologiste.

pénibilité du travail aurait un impact non seulement sur le niveau de  $\delta$  mais sur son taux d'accroissement. Leurs résultats empiriques montrent que l'état de santé des travailleurs manuels se détériore avec l'âge plus rapidement que celui des travailleurs non manuels.

Notre travail se situe dans la filiation de Case et Deaton. Le modèle présenté dans la section suivante prend explicitement en compte l'état de santé du salarié et repose sur l'idée que la pénibilité du travail détériore, à plus ou moins long terme, cet état de santé. Plus exactement, si les conditions de travail désignent des réalités très diverses de l'emploi, nous considérons exclusivement celles susceptibles d'agir sur la santé. La partie empirique de l'article porte ainsi sur une dimension des conditions de travail - l'irrégularité des horaires - dont l'effet sur la santé semble avéré, d'après les évidences empiriques. Nous nous intéressons donc à l'effet des conditions de travail sur les comportements d'absence pour maladie *via* l'état de santé du salarié. Ceci n'exclut évidemment pas que l'impact des conditions de travail sur les comportements puisse aussi passer par d'autres canaux, comme les phénomènes d'« aléa moral » décrits plus haut ou le simple fait que de mauvaises conditions de travail accroissent la désutilité de ce dernier. Mais l'identification de ces différents effets soulève des difficultés, sur lesquelles nous reviendrons dans la dernière section.



## II - Le cadre théorique

### II.1 Le modèle de base

Notre objectif ici est de comprendre, grâce à un modèle théorique, les mécanismes qui sont en jeu dans la décision de s'arrêter pour raison de santé et le rôle joué en la matière par les conditions de travail. Pour simplifier, nous supposons dans cette section que l'impact propre aux conditions de travail ne transite que par l'état de santé.

Nous partons de la modélisation proposée par Grossman (1972). Selon ce modèle, l'état de santé  $z_t$ , mesuré à la date  $t$  (i.e. à l'âge  $t$  de l'individu), est un bien qui se déprécie à un rythme  $\delta_t$ . Pour en ralentir la détérioration, l'individu se soigne, et ces soins « produisent » de l'état de santé en quantité  $i_t$ . L'évolution de  $z_t$  obéit donc à la relation suivante :

$$z_{t+1} = (1 - \delta_t)z_t + i_t \quad (1)$$

Nous supposons que le taux  $\delta_t$  dépend de caractéristiques d'environnement, et notamment des conditions de travail (Sickles et Taubman, 1986 ; Wagstaff, 1986). Il est plus élevé chez les travailleurs exerçant des travaux pénibles susceptibles d'altérer leur état de santé.

Concernant maintenant le lien entre santé et comportement d'absence, il y a *a priori* deux manières de le formaliser. La première est d'intégrer directement la dimension santé dans les paramètres de préférence du modèle de comportement, à la manière de Barmby *et alii* (1994). Une des difficultés de cette approche est qu'elle conduit à endogénéiser les préférences. Si, par exemple, l'individu décide de consacrer du temps à se soigner, son comportement doit en principe influencer sur son état de santé en l'améliorant, donc sur ses préférences à partir du moment où elles en dépendent. Pour cette raison, nous préférons suivre Case et Deaton (2005) et considérer l'état de santé comme un bien durable, qui s'use au cours du temps et que l'individu doit entretenir. Dans ce contexte, l'état de santé est un argument direct de la fonction d'utilité. Plus précisément, nous supposons que l'utilité dépend de la consommation  $c_t$  et de  $z_t$  :

$$u = u(c_t, z_t),$$

avec les hypothèses habituelles de concavité :  $u_c > 0$ ,  $u_z > 0$ ,  $u_{cc} < 0$ ,  $u_{zz} < 0$ . Nous supposons en outre que  $u$  est additivement séparable en ses arguments :  $u_{cz} = u_{zc} = 0$ .

Le comportement d'absence du salarié résulte alors d'un arbitrage mettant en jeu la consommation et la santé. Pour le voir, nous nous restreignons à une seule période, bien que notre modèle ait vocation à être inter-temporel. Soit  $\bar{h}$  la durée contractuelle de travail du salarié sur la période considérée. Si le salarié est malade, il va se soigner en s'absentant pendant une durée  $s$  et en consommant une quantité  $x$  de soins qu'il achète au prix  $\pi$ . En d'autres termes, la fonction de production d'état de santé dépend de deux arguments :  $i = i(s, x)$ , avec  $i_s > 0$  et  $i_x > 0$ . Bien que nous nous intéressions ici uniquement à l'absence pour maladie, nous enrichissons notre modèle en introduisant aussi la consommation de soins ou de médicaments<sup>3</sup>.

<sup>3</sup> Ceci devrait nous permettre, dans une phase ultérieure du travail et avec des données adaptées, d'étudier l'impact des conditions de travail sur la consommation de soins.

Soit  $z_0$  l'état de santé du salarié au début de la période. L'état de santé du salarié en fin de période vaut, d'après (1) :

$$z = (1 - \delta)z_0 + i(s, x) \quad (2)$$

Par ailleurs, la durée effective du travail du salarié sur la période est égale à  $\bar{h} - s$ . Soit  $w$  le taux de salaire et  $\tau$  le taux d'indemnisation des arrêts-maladie. Nous supposons, pour simplifier, que le salarié n'a pas d'autre source de revenus que son salaire et les indemnités de maladie qu'il perçoit en cas d'absence, salaires et indemnités que l'individu consomme totalement au cours de la période. Sa consommation au cours de la période est alors égale à :

$$c = w(\bar{h} - s) + \tau ws - \pi x = w\bar{h} - w(1 - \tau)s - \pi x \quad (3)$$

Dans ce cadre, la détermination de la durée optimale de l'absence résulte d'un arbitrage : toutes choses égales d'ailleurs, une durée d'absence  $s$  plus élevée améliore l'état de santé - expression (2) - et augmente l'utilité, mais dans le même temps diminue la consommation pour autant que  $\tau$  soit inférieur à 1 - expression (3) - ce qui génère une perte d'utilité.

Formellement, on a à l'optimum :

$$\frac{\partial}{\partial s} u(c, z) = 0 \text{ et } \frac{\partial}{\partial x} u(c, z) = 0$$

Nous retenons comme fonction d'investissement une Cobb-Douglas :  $i(s, x) = \theta s^\sigma x^{1-\sigma}$ , où  $\theta$  mesure l'efficacité des soins et  $\sigma$  le degré de substitution entre  $x$  et  $s$ . La double condition de l'optimum s'écrit :

$$\begin{cases} \frac{\partial}{\partial s} u(w\bar{h} - w(1 - \tau)s - \pi x, (1 - \delta)z_0 + \theta s^\sigma x^{1-\sigma}) = 0 \\ \frac{\partial}{\partial x} u(w\bar{h} - w(1 - \tau)s - \pi x, (1 - \delta)z_0 + \theta s^\sigma x^{1-\sigma}) = 0 \end{cases}$$

ou encore :

$$\begin{cases} w(1 - \tau)u_c(c, z) = \theta \sigma s^{\sigma-1} x^{1-\sigma} u_z(c, z) \\ \pi u_c(c, z) = \theta(1 - \sigma) s^\sigma x^{-\sigma} u_z(c, z) \end{cases} \quad (4)$$

La résolution de (4) fournit les durées  $s^*$  et  $x^*$  optimales. En divisant deux à deux les membres des deux équations de (4), on obtient :

$$\frac{x^*}{s^*} = w(1 - \tau) \frac{1 - \sigma}{\pi \sigma} \quad (5)$$

En posant  $\kappa = \theta \left( \frac{1 - \sigma}{\pi \sigma} \right)^{1-\sigma}$  et en utilisant (5), la première équation de (4) se réécrit :

$$[w(1 - \tau)]^\sigma u_c(c^*, z^*) = \kappa \sigma u_z(c^*, z^*) \quad (6)$$

avec :

$$\begin{cases} c^* = w\bar{h} - w\left(\frac{1-\tau}{\sigma}\right)s^* \\ z^* = (1-\delta)z_0 + \kappa[w(1-\tau)]^{1-\sigma}s^* \end{cases}$$

## II.2 L'introduction des conditions de travail dans le modèle

Nous introduisons maintenant les conditions de travail dans le modèle. Nous supposons au préalable que nous savons classer les emplois selon un indice  $p$  de pénibilité du travail, depuis les moins pénibles jusqu'aux plus pénibles. Nous supposons aussi que la pénibilité affecte le taux de détérioration de l'état de santé :

$$\delta = \delta(p), \text{ avec } \delta'(p) > 0.$$

Cette hypothèse, qui traduit l'idée d'une usure (temporelle) du capital santé par des conditions de travail défavorables, reprend les constats documentés dans la littérature de médecine ou d'ergonomie appliquées (voir la section précédente). Enfin, nous supposons que le salaire dépend aussi de  $p$  :

$$w = w(p), \text{ avec } w'(p) \geq 0.$$

Cette hypothèse s'appuie sur les pratiques salariales des entreprises, prévues par les conventions collectives, qui consistent à accorder aux salariés soumis à des conditions de travail *a priori* défavorables des suppléments de rémunération sous la forme de primes spécifiques (prime de nuit, de salissure, etc). Elle ne s'appuie pas expressément sur la théorie des différences compensatrices, et ne dépend donc pas de sa validité. Elle est toutefois renforcée si la théorie trouve des confirmations empiriques. Enfin, l'inégalité large ( $w'(p) \geq 0$ ) laisse ouverte la possibilité que les travaux pénibles ne soient pas spécifiquement rémunérés.

En considérant  $p$  comme exogène<sup>4</sup>, le théorème des fonctions implicites appliqué à (6) nous dit que le choix optimal  $s^*$  est fonction de  $p$  :  $s^* = s^*(p)$ . En dérivant (6) par rapport à  $p$ , on obtient, tous calculs faits :

$$\frac{\partial s^*}{\partial p} = w'(p)A + \delta'(p)B$$

avec :

$$A = \frac{[\sigma w^{\sigma-1}(1-\tau)^\sigma u_c + w^{\sigma-1}(1-\tau)^\sigma c^* u_{cc} - \kappa^2 \sigma (1-\sigma) w^{-\sigma} (1-\tau)^{1-\sigma} s^* u_{zz}]}{\left[ \frac{w^{\sigma+1}(1-\tau)^{\sigma+1}}{\sigma} u_{cc} + \kappa^2 \sigma w^{1-\sigma} (1-\tau)^{1-\sigma} u_{zz} \right]}$$

et :

$$B = \frac{\kappa \sigma z_0 u_{zz}}{\left[ \frac{w^{\sigma+1}(1-\tau)^{\sigma+1}}{\sigma} u_{cc} + \kappa^2 \sigma w^{1-\sigma} (1-\tau)^{1-\sigma} u_{zz} \right]}$$

où  $u_{cc} = u_{cc}(c^*, z^*)$  et  $u_{zz} = u_{zz}(c^*, z^*)$ . D'après les hypothèses posées sur les dérivées partielles de la fonction d'utilité, le dénominateur de  $A$  et  $B$  est négatif. Le

<sup>4</sup> Nous reviendrons plus loin sur cette hypothèse forte.

numérateur de  $A$  est positif à condition que l'utilité ne soit pas trop concave en  $c$  (c'est-à-dire que  $u_{cc}$  reste faible). Le numérateur de  $B$  est négatif. Dans ces conditions, si la pénibilité est compensée même partiellement par un supplément de salaire ( $w'(p) > 0$ ), alors son impact sur l'absence est ambigu puisqu'il résulte de deux mécanismes antagonistes :

- un effet incitatif au travail ou désincitatif à l'absence, lié au fait qu'un travail plus pénible offre aussi une rémunération plus avantageuse ;
- un effet incitatif à l'absence, lié au souci du salarié de préserver son état de santé.

Dans ces conditions, déterminer lequel de ces deux effets l'emporte devient une question empirique.

Notons pour terminer que si on considère  $\bar{h}$  et  $\tau$  comme des paramètres exogènes, en appliquant le même raisonnement qu'avec  $p$ , on obtient :

$$\left[ \frac{w^{\sigma+1}(1-\tau)^{\sigma+1}}{\sigma} u_{cc} + \kappa^2 \sigma w^{1-\sigma} (1-\tau)^{1-\sigma} u_{zz} \right] \frac{\partial s^*}{\partial \bar{h}} = w^{\sigma+1} (1-\tau)^{\sigma} u_{cc} ,$$

et :

$$\left[ \frac{w^{\sigma+1}(1-\tau)^{\sigma+1}}{\sigma} u_{cc} + \kappa^2 \sigma w^{1-\sigma} (1-\tau)^{1-\sigma} u_{zz} \right] \frac{\partial s^*}{\partial \tau} = \frac{w^{\sigma+1}(1-\tau)^{\sigma}}{\sigma} s^* u_{cc} - w^{\sigma} (1-\tau)^{\sigma-1} u_c .$$

On retrouve les résultats classiques établis dans la littérature économique (voir section 1), selon lesquels la propension à s'absenter augmente avec la durée contractuelle du travail ( $\partial s^* / \partial \bar{h} > 0$ ) et augmente avec le taux de compensation du salaire ( $\partial s^* / \partial \tau > 0$ ).

### III - Santé et conditions de travail : les problèmes empiriques

Sur le plan empirique, l'étude du rôle des conditions de travail sur l'absence pour maladie se heurte à au moins deux écueils. Le premier est la difficulté d'objectiver et de mesurer les conditions de travail. Le second a trait à la disponibilité de données adaptées, qui exige de fait une collecte relativement lourde. Ces deux problèmes nous ont contraints à faire des choix, pas toujours satisfaisants, sur les données utilisées et la variable de conditions de travail retenue.

#### ***III.1 Recueillir des données sur les conditions de travail et la santé : un processus de collecte exigeant***

La difficulté à objectiver les conditions de travail vient du fait que l'information recueillie en interrogeant les salariés sur leurs conditions de travail traduit à la fois la réalité de leur travail et la perception qu'ils en ont. Pour le montrer, Gollac (1997) a comparé les réponses fournies à deux enquêtes successives sur les conditions de travail, complémentaires aux enquêtes Emploi de l'INSEE. L'exemple emblématique, cité par Gollac, de la difficulté à séparer réalité et perception porte sur le tabagisme ambiant. Le pourcentage des employés administratifs déclarant respirer des fumées s'élève à 11 % en 1984 et à 21 % en 1991. Cette forte évolution ne reflète pas tant la dégradation des lieux de travail que la plus grande sensibilité des personnes à la présence de ces fumées, sensibilité qui a été aiguïlée par les campagnes des milieux médicaux, relayés par les pouvoirs publics, sur les méfaits du tabagisme. Dans le même ordre d'idées, Molinié (2003) a testé la cohérence des réponses fournies par les salariés aux questions sur leurs conditions de travail, en exploitant deux vagues successives de l'enquête *Santé, Travail et Vieillesse* (ESTEV)<sup>5</sup>, qui a interrogé les mêmes salariés à 5 ans d'intervalle. Par exemple, 24 %<sup>6</sup> des personnes enquêtées qui, en 1990, avaient déclaré travailler ou avoir travaillé en horaires alternants, ont, cinq ans plus tard, affirmé qu'elles ne l'avaient jamais fait. Le pourcentage est supérieur lorsqu'il s'agit de conditions de travail plus difficilement objectivables. Ainsi, avec le port de charges lourdes, il s'élève à 54 %.

En conséquence, une enquête ayant pour objectif de mesurer des conditions de travail afin d'en établir le lien avec tel ou tel comportement sur le marché du travail doit obéir à un protocole rigoureux. Le problème se complique lorsqu'on souhaite faire intervenir l'état de santé du salarié comme élément d'une chaîne causale reliant les conditions de travail au comportement observé. A moins de mettre en place un dispositif de collecte d'informations reposant sur des diagnostics objectifs, l'état de santé est auto-déclaré, en réponse à des questions de nature plus ou moins subjective. La variable pose donc des problèmes analogues à ceux soulevés par les indicateurs de conditions de travail. Bien plus, ces problèmes se cumulent au sens où, par exemple, les individus se sentant et se déclarant en mauvaise santé sont plus facilement enclins à « noircir » le tableau de leur environnement de travail, ce qui risque de biaiser toute corrélation établie empiriquement entre la santé et les conditions de travail. A ceci s'ajoute le fait que les effets des conditions de travail sur la santé, par exemple les symptômes consécutifs à une exposition prolongée à certains risques, peuvent mettre du temps à se manifester. Les données doivent donc tenir compte de cette dimension dynamique inhérente aux phénomènes étudiés. En d'autres termes, et à cause des problèmes de mémoire comme ceux mis en évidence

---

<sup>5</sup> L'enquête ESTEV, coordonnée par l'Inserm et le Creapt (Centre de Recherche et d'Etude sur l'Age et la Population au Travail), a été menée par près de 400 médecins du travail volontaires auprès d'un échantillon de salariés qu'ils suivaient, sélectionné sur leur année de naissance (1938, 1943, 1948, 1953). Environ 21 000 salariés ont répondu à l'enquête de 1990, et 87 % d'entre eux à l'enquête de 1995.

<sup>6</sup> Calcul à partir des données figurant dans le tableau annexé à l'article.



par Molinié (voir *supra*), l'information doit être recueillie en interrogeant les mêmes individus à plusieurs reprises, sur une période suffisamment longue, et en limitant les questionnements rétrospectifs.

### **III.2 La source retenue**

A notre connaissance, il n'existe pas de source de données qui réponde de manière satisfaisante aux exigences que nous avons détaillées dans la section précédente. Les quelques panels disponibles ne contiennent pas d'information suffisamment complète et précise à la fois sur les conditions de travail et les autres caractéristiques de l'emploi occupé, sur l'état de santé, et sur les arrêts maladie. Par ailleurs, nous n'avons pas non plus de données sur le taux de couverture des indemnités journalières (le paramètre  $\tau$  du cadre théorique), qui peut varier substantiellement d'une entreprise à l'autre (voir encadré). La conséquence est donc qu'il n'est pas possible empiriquement d'identifier séparément l'effet direct sur la santé et les comportements d'arrêts des salariés, à niveau de santé donné.

#### **L'indemnisation des congés maladie en France**

D'après le code de la Sécurité Sociale et le droit du travail, un salarié peut s'absenter de son travail pour maladie, à condition d'adresser dans un délai de 48 heures un certificat médical à son employeur. Le salarié qui ne le fait pas peut se voir infliger une sanction disciplinaire, voire être licencié pour faute.

En principe, la maladie entraîne une simple suspension du contrat de travail : l'employeur doit donc réintégrer le salarié à son poste de travail à l'issue de son congé. Toutefois, si les absences prolongées ou répétées du salarié désorganisent la bonne marche de l'entreprise et rendent nécessaire le remplacement définitif du salarié malade, l'employeur peut le licencier.

Pendant son arrêt maladie, un salarié du secteur privé peut percevoir plusieurs types de compensation salariale. En premier lieu, des indemnités journalières sont versées par la Sécurité Sociale à partir du 4<sup>ème</sup> jour d'arrêt de travail. Il existe donc un délai de carence de 3 jours. Les indemnités journalières sont égales à 50% du salaire brut, dans la limite du plafond de Sécurité sociale (2 352 €/mois pour 2002) pendant les 6 premiers mois et à 51,49 % au-delà. Ces indemnités peuvent être majorées si le salarié a au moins 3 enfants à charge. A noter également que, pour recevoir ces indemnités, le salarié doit justifier d'un montant suffisant de cotisations dans le passé (au moins 1015 fois le SMIC horaire dans les six derniers mois pour les arrêts de moins de 6 mois).

Ces indemnités peuvent être complétées par l'employeur, selon des conditions fixées par la convention collective, l'accord d'entreprise ou l'accord de mensualisation. Certaines conventions prévoient en particulier le versement d'indemnités journalières durant les trois jours de carence. L'ancienneté du salarié dans l'entreprise est un paramètre important dans les modes d'indemnisation.

Enfin, des mutuelles ou des assurances privées peuvent également verser un complément. Ces assurances peuvent être directement souscrites par le salarié, proposées ou imposées par l'entreprise. La première enquête réalisée sur le sujet montre une extrême diversité des situations selon les entreprises (Couffignal *et alii*, 2004). Près des deux tiers des salariés du secteur privé disposeraient d'une mutuelle complémentaire.

Notre travail empirique a finalement été effectué sur les données de l'enquête trimestrielle sur l'emploi de l'Insee. Certes, elle ne possède pas toutes les informations

qu'on aurait souhaitées. Ainsi, elle en fournit sur les conditions de travail qui restent parcellaires. Quelques variables décrivent l'organisation des horaires du salarié, mais on ne sait rien sur les efforts ou les risques au travail, ou bien sur la charge mentale. Autre inconvénient de taille, on ignore tout de l'état de santé de l'enquêté. L'absence pour maladie n'est mesurée que sur une seule semaine, dite de référence, qui est en général la semaine calendaire précédant le jour de l'interrogation. Enfin, la dimension de panel de l'enquête n'est pas réellement utilisable pour notre propos, parce que sa profondeur de 15 mois est insuffisante pour traiter correctement les aspects dynamiques du problème qui nous occupe et parce que toutes les questions de l'enquête ne sont pas posées à chaque vague d'interrogation.

Néanmoins, l'enquête trimestrielle sur l'emploi a plusieurs atouts. La variable de conditions de travail que nous avons retenue a l'avantage sur d'autres d'être plus simplement objectivable. Elle est liée aux rythmes de travail, et distingue trois catégories de salariés : ceux dont les horaires sont les mêmes d'une semaine à l'autre, ceux ayant des horaires alternés et ceux dont les horaires varient d'une semaine à l'autre. Autre argument, l'échantillon est de taille suffisante pour nous permettre de restreindre notre champ à une population homogène, en l'espèce les salariés masculins du secteur privé classés comme ouvriers. L'analyse aurait été probablement beaucoup plus délicate si on l'avait étendue aux différentes catégories de salariés, car cela aurait été la source d'une importante hétérogénéité inobservée. Par exemple, travailler en horaires alternés ne recouvre pas les mêmes réalités pour un ouvrier et pour un cadre. En outre, les ouvriers sont surreprésentés dans les absences pour maladie<sup>7</sup>. Enfin, l'enquête donne beaucoup d'informations sur le poste de travail (la fonction, la qualification, le salaire) et l'entreprise (le secteur d'activité, la taille), ce qui est important pour la méthode empirique utilisée (voir *infra*).

L'échantillon d'étude a été constitué de la manière suivante. Dans chacune des 12 enquêtes s'étalant entre le 1<sup>er</sup> trimestre 2002 et le 4<sup>ème</sup> trimestre 2004, nous avons sélectionné les ouvriers de sexe masculin, âgés de 18 à 59 ans, ayant un emploi dans le secteur privé<sup>8</sup> et qui répondaient pour la première fois à l'enquête. L'effectif est de 13 656 salariés. Environ 3,6% de ces individus ont déclaré avoir été absents toute la semaine de référence pour raison de maladie. Notons que l'on s'intéresse ici aux absences pour maladie durant au moins une semaine. On exclut donc de l'analyse les arrêts (très) courts.

Pour disposer d'un nombre plus important de personnes s'étant arrêtées pour maladie, et pour prendre en compte - même très partiellement - l'impact différé des conditions de travail sur la santé, nous avons recueilli aussi les réponses des salariés à la question sur l'absence qui leur a été posée lors de la deuxième interrogation 3 mois plus tard. Il a fallu toutefois restreindre l'échantillon, puisqu'on a dû exclure ceux qui n'avaient pas répondu à la deuxième interrogation et ceux qui, tout en ayant répondu, n'étaient plus en emploi. Ceci n'est d'ailleurs pas sans risque sur les estimations s'il se trouve que les personnes qui n'ont pas été retrouvées ou qui ont perdu leur emploi ont des caractéristiques, notamment d'état de santé (inobservé dans l'enquête), différentes des autres personnes. Nous reviendrons sur ce point ultérieurement. Si on comptabilise les réponses fournies par les salariés, lors de la deuxième interrogation, à la question sur leur absence pour maladie, alors la part de ceux qui se sont arrêtés (lors d'une des deux semaines de référence) s'établit à 5,8 %, sur un effectif total de 11 538 salariés.

Sur cet échantillon, nous avons regroupé les 21,4 % d'ouvriers travaillant en horaires alternants et les 14,8 % travaillant en horaires irréguliers d'une semaine sur l'autre.

---

<sup>7</sup> En 2004, 41% des arrêts maladies concernaient des ouvriers, qui ne représentent que 27% des effectifs salariés (point d'information de la CNAM-TS : <http://www.ameli.fr/pdf/1550.pdf>).

<sup>8</sup> A l'exception des secteurs de l'énergie, des activités financières et immobilières, de l'éducation, de la santé, de l'action sociale, dont les effectifs sont très faibles pour cette catégorie de salariés.

Cet ensemble constitue la catégorie des ouvriers en horaires irréguliers<sup>9</sup>, que nous opposons aux 63,8 % dont les horaires sont les mêmes d'une semaine sur l'autre.

En moyenne, le salaire reçu par les ouvriers travaillant en horaires irréguliers s'établit à 1 394 €, soit 143 € de plus que celui perçu par ceux ayant des horaires réguliers. La différence subsiste lorsqu'on la contrôle par les caractéristiques du salarié, de son poste de travail et de son entreprise. Ce constat est tout à fait cohérent avec les résultats de Lanfranchi *et alii* (2002) qui, sur données françaises, ont estimé à 16 % le différentiel de salaire entre les salariés en horaires décalés et les autres.

Parallèlement, quelques travaux étrangers ont documenté l'impact négatif de l'irrégularité des horaires sur la santé. Spécifiquement et par exemple, le travail décalé ou de nuit perturbe les cycles de sommeil et les habitudes alimentaires, ce qui dans certains cas conduit à des désordres du système intestinal augmentant les risques de cancer, ou du système neuropsychique favorisant les états dépressifs (Costa, 1996).

Tout compte fait, dans le cas présent où la variable de conditions de travail est le type d'horaires, on se retrouve dans la situation où leur impact sur les absences pour maladie est théoriquement ambigu (puisque  $w'(p) > 0$  et  $\delta'(p) > 0$  : voir section 2.2). Ce sont donc les données qui nous permettent de trancher.

### ***III.3 Quelques statistiques sur l'irrégularité des horaires et l'absence pour maladie***

La table 1 présente les principales caractéristiques de l'échantillon. Il distingue les salariés en horaires irréguliers des autres.

A première vue, les caractéristiques d'entreprise sont beaucoup plus discriminantes que les caractéristiques individuelles. Notamment, la taille joue un grand rôle : les postes de travail à horaires irréguliers sont bien plus fréquents dans les grandes unités. La situation est aussi très contrastée selon les secteurs d'activité. Enfin, la régularité des horaires est liée à l'existence d'un accord de modulation ou d'annualisation du temps de travail dans l'entreprise. Les caractéristiques propres du salarié jouent moins, à l'exception de la catégorie socioprofessionnelle. Ainsi, les ouvriers exerçant une activité à caractère artisanal ont des horaires plus réguliers que les autres. Ces premières statistiques montrent l'importance d'avoir de nombreuses informations sur l'entreprise si on veut contrôler les différences observées entre les postes à horaires réguliers et ceux à horaires irréguliers<sup>10</sup>.

---

<sup>9</sup> Dans toute la suite et pour faire court, on regroupera sous la dénomination *horaires irréguliers* les horaires alternants et ceux variant d'une semaine à l'autre.

<sup>10</sup> C'est un des avantages de l'enquête Emploi, comparativement à d'autres sources concurrentes sur le sujet « conditions de travail et santé », comme l'enquête Santé qui contient très peu d'informations sur l'entreprise dans laquelle travaille l'individu enquêté.

Table 1 - Caractéristiques des salariés selon la régularité de leurs horaires

	Horaires irréguliers	Horaires réguliers
<i>Age (%)</i>	<i>100.0</i>	<i>100.0</i>
18-29 ans	25.6	25.5
30-39 ans	31.9	29.8
40-49 ans	26.2	26.5
50-59 ans	16.3	18.1
<i>Age de fin d'études (années)</i>	<i>17.7</i>	<i>17.4</i>
<i>Possession d'un CAP-BEP (%)</i>	<i>100.0</i>	<i>100.0</i>
non	56.1	53.8
oui	43.9	46.2
<i>Contrat à durée indéterminée (%)</i>	<i>100.0</i>	<i>100.0</i>
non	11.4	11.8
oui	88.6	88.2
<i>Ancienneté dans l'entreprise (%)</i>	<i>100.0</i>	<i>100.0</i>
moins d'un an	16.6	18.4
de 1 an à moins de 5 ans	27.9	30.6
de 5 ans à moins de 10 ans	15.1	14.0
10 ans ou plus	40.4	36.8
<i>Catégorie socioprofessionnelle (%)</i>	<i>100.0</i>	<i>100.0</i>
ouvriers qualifiés de type industriel	37.0	22.9
ouvriers qualifiés de type artisanal	10.0	32.1
chauffeurs	17.2	10.3
ouvriers qualifiés de la manutention ...	11.8	8.6
ouvriers non qualifiés de type industriel	18.3	13.0
ouvriers non qualifiés de type artisanal	3.2	9.8
ouvriers agricoles et assimilés	2.4	3.4
<i>Fonction exercée (%)</i>	<i>100.0</i>	<i>100.0</i>
production, chantier, exploitation	55.8	54.7
installation, réparation, maintenance	10.9	18.0
nettoyage, logistique, ...	23.8	20.3
autres fonctions	9.5	7.0
<i>Secteur d'activité (%)</i>	<i>100.0</i>	<i>100.0</i>
agriculture, sylviculture, pêche	2.4	4.3
industries agricoles et alimentaires	8.2	5.3
industries des biens de consommation	5.5	3.6
industrie automobile	8.3	2.9
industries des biens d'équipement	6.9	7.8
industries des biens intermédiaires	26.9	13.1
construction	6.4	28.5
commerce et réparations	8.8	17.1
transports	19.3	8.1
services aux entreprises	5.0	5.8
services aux particuliers	2.4	3.5
<i>Taille de l'entreprise (%)</i>	<i>100.0</i>	<i>100.0</i>
moins de 10 salariés	11.6	29.5
de 10 à 49 salariés	21.2	35.5
de 50 à 199 salariés	23.9	19.1
200 salariés ou plus	41.0	13.4
inconnue	2.3	2.5
<i>Accord de modulation ou d'annua- lisation du temps de travail (%)</i>	<i>100.0</i>	<i>100.0</i>
non	71.7	83.1
oui	28.3	16.9
<i>Grande région de résidence (%)</i>	<i>100.0</i>	<i>100.0</i>
Ile-de-France	7.6	12.2
Centre	27.7	23.8
Nord	14.9	8.6
Est	13.3	11.2
Ouest	12.9	13.8
Sud Ouest	8.3	10.3
Rhône-Alpes et Auvergne	10.4	10.8
Sud Est	5.0	9.3
Nombre d'observations	4 184	7 354

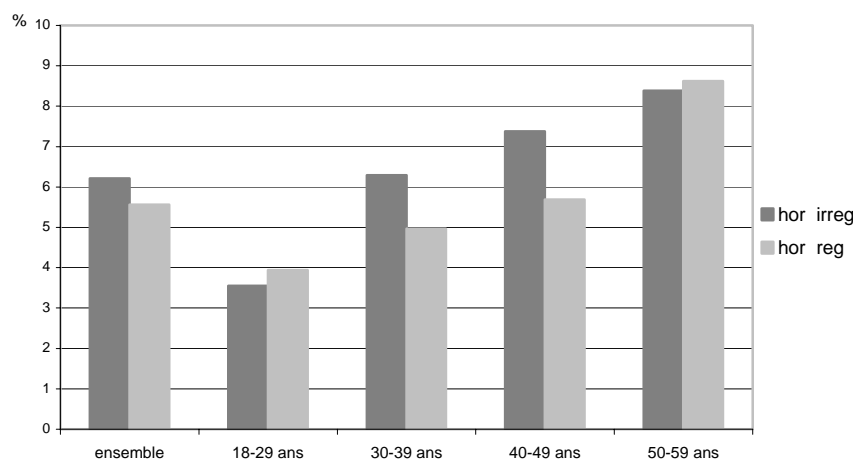
Champ : ouvriers du secteur privé, âgés de 18 à 59 ans

Source : enquêtes Emploi en continu 1<sup>er</sup> trim. 2002 - 4<sup>ème</sup> trim. 2004.

La figure 1 représente la part des ouvriers s'étant absentés pour maladie selon la régularité de leurs horaires de travail. Tous âges confondus, les salariés en horaires irréguliers se sont arrêtés un peu plus souvent (taux d'absence de 6,2 %) que les autres (taux d'absence de 5,6 %). Toutefois, un test de comparaison des deux taux ne rejette pas l'hypothèse nulle de leur égalité.

Sans surprise, le taux d'absence augmente avec l'âge du salarié. De manière plus intéressante, sa relation avec la régularité des horaires ne va pas toujours dans le même sens. Apparemment, les plus jeunes comme les plus âgés travaillant en horaires irréguliers s'arrêteraient un peu moins souvent que les salariés des tranches d'âge correspondantes occupant des postes à horaires réguliers. Ce constat purement empirique confirme l'absence, établie par le modèle théorique, de lien clair entre conditions de travail et arrêt-maladie.

**Figure 1 - Taux d'arrêts maladie selon la régularité des horaires de travail**



Lecture : 6,2 % des ouvriers en horaires irréguliers se sont arrêtés pour maladie

Champ : ouvriers du secteur privé, âgés de 18 à 59 ans

Source : enquêtes Emploi en continu 1<sup>er</sup> trim. 2002 - 4<sup>ème</sup> trim. 2004.

Toutefois, les différences dans les taux d'absence révélées par la figure 1 sont peut-être dues à d'autres facteurs que l'irrégularité des horaires. On a vu, en effet, que les postes à horaires irréguliers se trouvent plus fréquemment dans les entreprises industrielles (table 1). Or les risques professionnels y sont plus importants. Ils pourraient être, davantage que l'irrégularité des horaires, la cause des absences pour maladie.

## IV - Méthode économétrique

### IV.1 Méthode d'appariement

Pour dépasser ces constats empiriques probablement biaisés, nous avons cherché à contrôler l'hétérogénéité observée entre les salariés en horaires irréguliers et les salariés en horaires réguliers, hétérogénéité due pour beaucoup au fait que les postes de travail ne se situent pas dans les mêmes entreprises. Pour ce faire, nous avons utilisé des méthodes d'appariement (*matching*). En deux mots, leur principe est le suivant<sup>11</sup>. On associe d'abord à chaque salarié travaillant en horaires irréguliers un « jumeau » - un salarié ayant les mêmes caractéristiques (observables) que lui - qui travaille en horaires réguliers. Puis on compare les absences pour maladie des salariés en horaires irréguliers et de leurs « jumeaux ».

#### IV.1.1 Principe et hypothèses de la méthode

La méthode peut être formalisée ainsi. Soit  $I$  le type d'horaires de travail des salariés, avec  $I = 1$  lorsque l'individu travaille en horaires irréguliers, et  $I = 0$  sinon. Soit  $M$  la variable binaire mesurant l'absence pour maladie :  $M = 1$  si le salarié a déclaré s'être arrêté,  $M = 0$  sinon. Le principe de la méthode consiste à systématiquement définir pour chaque salarié, qu'il soit ou non en horaires irréguliers, deux « variables de résultat » :  $M(1)$  mesurant l'absence pour maladie s'il travaille ou s'il travaillait en horaires irréguliers, et  $M(0)$  mesurant l'absence pour maladie s'il travaille ou s'il travaillait en horaires réguliers. Bien entendu, pour chaque salarié, une seule de ces deux variables est observée :  $M(1)$  pour celui qui, de fait, travaille en horaires irréguliers ( $I = 1$ ),  $M(0)$  pour celui qui travaille en horaires réguliers ( $I = 0$ ).

Ce que l'on cherche à estimer ici est l'impact (moyen) que le seul fait de travailler en horaires irréguliers a sur l'absence pour maladie. Formellement, cet impact s'écrit :

$$\Delta = E(M(1) - M(0) | I = 1)$$

Il est estimé sur les seuls salariés en horaires irréguliers ( $I = 1$ ), et est égal à la différence, en moyenne, entre ce qu'ont été réellement leurs absences pour maladie ( $M(1)$ ) et ce qu'elles auraient été s'ils avaient travaillé dans des postes équivalents mais à horaires réguliers ( $M(0)$ ). La difficulté est, encore une fois, que seule  $M(1)$  est observée chez ces salariés. Sans hypothèse supplémentaire, on ne peut donc pas estimer  $\Delta$ .

L'hypothèse - identifiante - que l'on pose est la suivante : *conditionnellement* à un ensemble de caractéristiques observées  $X$ , le fait d'être absent pour maladie en travaillant en horaires réguliers (c'est-à-dire la quantité  $M(0)$ ) n'est pas lié au fait d'avoir choisi de travailler en horaires réguliers ou en horaires irréguliers. Formellement, l'hypothèse s'écrit :

$$M(0) \perp I \mid X \tag{7}$$

En clair, prenons un individu, de caractéristiques  $X$ , qui travaille en horaires irréguliers. Si ses horaires étaient réguliers, il se comporterait de la même manière qu'un individu ayant les mêmes caractéristiques  $X$  et qui travaille effectivement en horaires réguliers. En toute rigueur, l'hypothèse suivante, moins forte, suffit :

---

<sup>11</sup> Pour une présentation récente de ces méthodes et une discussion de leurs avantages et inconvénients, voir Smith et Todd (2005).

$$E(M(0)|X, I=1) = E(M(0)|X, I=0) \quad (8)$$

Avec cette hypothèse, on peut maintenant estimer  $\Delta$ . En effet, la loi des espérances itérées nous permet d'écrire :

$$E(M(0)|I=1) = E(E(M(0)|X, I=1)|I=1) = E(E(M(0)|X, I=0)|I=1),$$

si bien que :

$$\Delta = E(M(1) - M(0)|I=1) = E(M(1) - E(M(0)|X, I=0)|I=1),$$

c'est-à-dire :

$$\Delta = E(M - E(M|X, I=0)|I=1),$$

où  $M$  sont maintenant les absences *observées* (i.e :  $M = M(0)$  si le salarié travaille effectivement en horaires réguliers,  $M = M(1)$  sinon). Pour estimer  $\Delta$ , la démarche est alors la suivante :

- pour chaque salarié  $i$  de caractéristiques  $X_i$  travaillant en horaires irréguliers (dont on observe l'absence  $M_i$ ), on retient le(s) salarié(s) en horaires réguliers ayant les mêmes caractéristiques  $X_i$  ;
- on estime la moyenne  $E(M|X_i, I=0)$  des absences de ces « jumeaux » ; cette quantité est appelée « valeur contrefactuelle » - ou plus simplement « contrefactuel »<sup>12</sup> - de la valeur observée  $M_i$  ;
- on en déduit l'écart  $M_i - E(M|X_i, I=0)$  ;
- on calcule la moyenne de ces écarts sur le sous-échantillon des travailleurs en horaires irréguliers.

#### IV.1.2 Le score de propension

Pour réduire le biais d'hétérogénéité, il vaut mieux faire l'appariement sur de nombreuses caractéristiques  $X$ . Mais si elles le sont trop et si plusieurs d'entre elles sont continues, alors il est fort possible qu'on ne sache pas mettre en face d'un salarié travaillant en horaires irréguliers, un salarié en horaires réguliers ayant *strictement* les mêmes caractéristiques. Et si cela se produit fréquemment, alors on ne peut pas espérer estimer une « bonne » valeur de  $\Delta$ .

Rosenbaum et Rubin (1983) ont proposé un moyen de résoudre le problème, qui conduit à la méthode d'*appariement sur le score de propension* (*propensity score matching*). Ils s'intéressent de manière générale aux fonctions  $b(X)$  des caractéristiques  $X$  vérifiant la propriété suivante (*balancing property* dans la littérature anglo-saxonne) :

$$X \perp I \mid b(X) \quad (9)$$

Exprimée littéralement, cette propriété signifie que, conditionnellement à l'index synthétique  $b(X)$ , les personnes qui travaillent en horaires réguliers sont bien identiques (en termes de caractéristiques observables  $X$ ) aux personnes qui travaillent en horaires irréguliers. Rosenbaum et Rubin ont montré que, pour toute fonction  $b(X)$  vérifiant la propriété (9), si  $M(0)$  est indépendant de  $I$  conditionnellement à  $X$  (i.e. si (7) est vraie), alors  $M(0)$  l'est conditionnellement à la fonction  $b(X)$  :

<sup>12</sup> Littéralement : à l'opposé du fait (observé).

$$M(0) \perp I \mid X \Rightarrow M(0) \perp I \mid b(X) \quad (10)$$

La propriété (9) est vérifiée pour une fonction particulière de  $X$ , appelée score de propension (*propensity score*), et qui est égale à la probabilité de travailler avec des horaires irréguliers  $\pi(X) = \Pr(I=1 \mid X)$ . Par conséquent, si l'hypothèse identifiante (7) est vérifiée, alors :

$$M(0) \perp I \mid \pi(X)$$

Dans ces conditions, on peut formellement reprendre la démarche d'estimation de  $\Delta$  détaillée en 4.1.a, en remplaçant  $X_i$  par  $\pi(X_i)$ , la probabilité de travailler en horaires irréguliers. L'appariement se fait non plus sur les  $X$  mais sur le score de propension  $\pi(X)$  : on associe à chaque salarié en horaires irréguliers un salarié en horaires réguliers dont le score est le plus proche possible.

Le problème est qu'on a remplacé des grandeurs observées  $X$  par une variable  $\pi(X)$  qui ne l'est pas. L'appariement doit donc se faire non pas sur le score de propension  $\pi(X)$  mais sur une valeur estimée  $\hat{\pi}(X)$  du score. Il faut que cette fonction  $\hat{\pi}(X)$  vérifie la propriété (9) de manière à pouvoir adapter la démarche d'estimation de  $\Delta$  (section 4.1.a) en remplaçant  $X$  par  $\hat{\pi}(X)$ . Concrètement, on obtient  $\hat{\pi}(X)$  en estimant un modèle *logit* « expliquant » le type d'horaires  $I$  par les caractéristiques  $X$  :

$$\pi(X) = \frac{e^{h(X)}}{1 + e^{h(X)}}$$

où  $h(X)$  est un polynôme en  $X$  de degré 1 ou plus. On détermine alors  $h$ , c'est-à-dire les variables à élever au carré (ou au cube ...) et les termes d'interaction à introduire, de telle manière que la propriété (9) soit vérifiée pour  $\hat{\pi}(X)$ .

Reste à trouver un test convaincant de cette propriété. A notre connaissance, il n'existe pas actuellement de consensus sur ce sujet<sup>13</sup>. Les propriétés statistiques des tests utilisés dans la littérature ne sont pas clairement établies. Dans la mesure où il semble le plus répandu, et pour sa simplicité, nous avons retenu ici un test d'égalité des moyennes : on compare la valeur moyenne de chaque variable sur le sous-échantillon des personnes qui travaillent avec des horaires irréguliers avec la valeur moyenne de cette variable sur leur(s) « jumeau(x) », c'est à dire les personnes qui travaillent avec des horaires réguliers et ont un score estimé proche.

On commence donc par spécifier le logit du score de manière parcimonieuse (polynôme de degré 1). On teste l'égalité des moyennes de chaque variable  $X$  calculées sur le groupe des personnes en horaires irréguliers et sur le groupe des personnes en horaires réguliers. Si le test n'est pas vérifié pour l'une ou l'autre des variables  $X$ , il faut modifier la spécification du logit en introduisant des termes d'ordre 2 et répéter la procédure avec cette nouvelle spécification.

#### IV.1.3 L'estimation du contrefactuel

Une fois le score estimé pour tous les individus de l'échantillon, on passe à l'estimation du contrefactuel  $E(M \mid \hat{\pi}(X_i), I=0)$ . Il faut auparavant s'assurer que pour chaque individu en horaires irréguliers, on puisse trouver au moins un individu en horaires réguliers ayant le même score et qui puisse donc lui être apparié. Il faut donc

<sup>13</sup> Comme l'illustre la controverse entre Smith and Todd d'une part, et Dehejia de l'autre, dans le numéro 125(1-2) du Journal of Econometrics (2005).



que les deux distributions du score (des salariés en horaires réguliers d'une part, et des salariés en horaires irréguliers d'autre part) aient un *support commun* qui soit suffisamment large. En pratique, nous avons décidé d'éliminer les salariés en horaires irréguliers dont le score de propension est supérieur à la valeur maximale du score des salariés en horaires réguliers. Cette manière simple de faire est ici facilitée notamment par la taille importante de notre échantillon<sup>14</sup>.

Il y a plusieurs possibilités d'estimer le contrefactuel (voir Caliendo et Kopeinig, 2005, pour une description détaillée). La plus immédiate est de retenir, pour chaque salarié en horaires réguliers, son plus proche voisin, c'est-à-dire le salarié en horaires irréguliers qui a le score le plus proche du sien. Le contrefactuel est simplement la valeur prise par  $M$  pour ce plus proche voisin. La simplicité de la méthode a cependant des revers : si le « plus proche voisin » d'un salarié en horaires irréguliers est en fait très éloigné, le contrefactuel risque d'être biaisé ; à l'inverse, si pour un salarié en horaires irréguliers il existe de nombreux salariés comparables en terme de propension, il est dommageable (et inefficace) de ne retenir que l'un d'entre eux. L'estimateur est donc souvent peu précis. Pour pallier ces deux problèmes, les solutions consistent à 1) ne conserver que les voisins « suffisamment proches », 2) retenir plusieurs « plus proches voisins ».

Une autre possibilité est d'utiliser l'estimateur à noyau, proposé par Heckman *et alii* (1998). C'est celle que nous avons choisie. Elle consiste à retenir, pour chaque salarié en horaires irréguliers, tous les salariés en horaires réguliers, mais en les affectant d'un poids inversement proportionnel à leur « distance » avec le salarié en horaires irréguliers (c'est-à-dire à la différence des scores). Ces poids sont des noyaux de densité. Le contrefactuel est estimé par :

$$\hat{E}[M | \hat{\pi}(X_i), I = 0] = \sum_{j \in \{I=0\}} w_{ij} M_j \quad (11)$$

$$\text{avec : } w_{ij} = \frac{K[(\hat{\pi}(X_j) - \hat{\pi}(X_i))/h]}{\sum_{j \in \{I=0\}} K[(\hat{\pi}(X_j) - \hat{\pi}(X_i))/h]} \quad (12)$$

où  $h$  est la fenêtre du noyau  $K$ . Parmi les noyaux possibles, nous avons opté pour le noyau normal :

$$K(z) = \phi(z),$$

où  $\phi$  est la fonction de densité de la loi normale centrée réduite. Quant à la fenêtre, on retient :

$$h = \sigma_{\hat{\pi}(X)} \cdot n_0^{-\frac{1}{5}},$$

où  $\sigma_{\hat{\pi}(X)}$  est l'écart-type du score de propension estimé, calculé sur le sous-échantillon (de taille  $n_0$ ) des salariés en horaires réguliers (et dont le score de propension appartient au support commun).

Cet estimateur est *a priori* plus précis (il utilise davantage d'informations), mais moins performant en termes de correction du biais (tous les salariés en horaires réguliers interviennent, certes avec une pondération d'autant plus faible qu'ils sont éloignés). En pratique, nous avons vérifié que nos résultats variaient peu avec le choix de la méthode retenue (l'estimateur du plus proche voisin « simple » étant - logiquement -

<sup>14</sup> Dans des cas moins favorables, la détermination du support peut être plus délicate (voir par exemple Caliendo et Kopeinig, 2005).

le plus sensible à la spécification du score de propension retenue). Enfin, pour calculer l'écart-type de  $\Delta$ , nous avons utilisé le bootstrap.

#### IV.1.4 Le choix des variables explicatives

La validité de la méthode demande de disposer d'un ensemble de variables observables  $X$  suffisamment complet pour que l'hypothèse (7) soit vérifiée. Ces variables observables doivent être des déterminants *à la fois* du fait de travailler en horaires irréguliers, et de la probabilité d'être absent pour maladie.

Nous avons retenu ici des variables individuelles « classiques », telles l'âge, le diplôme (disposer d'un CAP ou d'un BEP), l'âge de fin d'études et la nationalité, dans la mesure où ces caractéristiques peuvent être corrélées avec des choix professionnels importants et avec l'état de santé de la personne.

La richesse de l'enquête Emploi permet également de contrôler les estimations par des variables d'entreprise : secteur d'activité de l'établissement, taille de l'entreprise, présence d'un accord de réduction de temps de travail. Nous avons également utilisé des caractéristiques de l'emploi occupé (la catégorie socioprofessionnelle) ainsi que le statut du contrat (en CDI) et l'ancienneté dans l'emploi : ces variables, qui affectent le degré de sécurité dans l'emploi ainsi que le taux de compensation des arrêts maladie, sont en effet susceptibles d'intervenir dans le choix de s'arrêter pour maladie.

Nous ne pouvons mesurer ici que des effets moyens des horaires irréguliers sur l'absentéisme pour maladie. Cependant, il est probable que ces effets soient très hétérogènes. L'âge est une dimension particulièrement déterminante : avec le temps, le « capital santé » diminue, orientant les choix d'activité, et l'horizon temporel n'est pas le même. Nous avons donc effectué des estimations séparées par groupe d'âge.

#### IV.1.5 En guise de résumé

En résumé, la méthode d'estimation par appariement sur le score de propension se déroule de la manière suivante :

Etape 1 : On estime un modèle logit expliquant l'irrégularité des horaires par les variables explicatives retenues ; on en déduit, pour chaque individu de l'échantillon, la valeur de son score, c'est-à-dire la probabilité d'être en horaires irréguliers prédite par le logit compte tenu des caractéristiques de l'individu.

Etape 2 : On sélectionne les individus de l'échantillon appartenant au support commun : on élimine les ouvriers en horaires irréguliers dont le score excède la valeur maximale du score chez les salariés en horaires réguliers.

Etape 3 : Pour chaque salarié en horaire irrégulier, on estime le contrefactuel par la méthode du noyau (expression (11) *supra*)<sup>15</sup> ; on en déduit l'écart du contrefactuel à la valeur observée ; l'impact de l'irrégularité des horaires sur l'absence pour maladie est ensuite estimé par la moyenne empirique de ces écarts calculée sur l'ensemble des ouvriers en horaires irréguliers.

---

<sup>15</sup> Les estimations ont été menées avec la procédure *psmatch2* (Leuven et Sianesi, 2005) du logiciel Stata. On trouvera en annexe les programmes qui ont été écrits pour les estimations avec le logiciel sas.

**Etape 4 :** On teste la propriété (9) de la fonction de score. Pour ce faire, on procède comme suit. Pour chaque variable explicative  $x$  de la fonction de score :

- on calcule sa moyenne empirique sur le sous-échantillon des ouvriers en horaires irréguliers ;
- pour chaque ouvrier en horaire irrégulier, on calcule sa valeur « contrefactuelle » :

$$\tilde{x}_i = \sum_j w_{ij} \cdot x_j$$

où  $w_{ij}$  sont les poids donnés par (12) ; on calcule ensuite la moyenne empirique de ces valeurs contrefactuelles sur le sous-échantillon des ouvriers en horaires irréguliers ;

- on teste l'égalité de ces deux moyennes ; si on ne peut rejeter l'hypothèse nulle d'égalité des moyennes pour aucune des variables explicatives, alors la fonction de score estimée par le logit possède bien la propriété (9) ; sinon, la spécification du logit est reprise en croisant la variable explicative  $x$ , pour laquelle l'hypothèse nulle d'égalité des moyennes est rejetée, avec d'autres.

**Etape 5 :** Une fois la bonne spécification du logit obtenue (*i.e.* qui passe le test de la *balancing property*), on retient comme valeur estimée de l'impact des horaires irréguliers sur l'absence pour maladie celle calculée lors de l'étape 3, et on estime son écart-type par bootstrap.

#### ***IV.2 Une estimation alternative de l'impact des horaires irréguliers***

La méthode par appariement repose crucialement sur l'hypothèse (7). Même en contrôlant par de nombreuses caractéristiques  $X$ , il y a toujours le risque que des variables inobservées ou inobservables susceptibles d'influer sur  $M$  soient distribuées différemment chez les salariés en horaires irréguliers et chez ceux en horaires réguliers. L'état de santé du salarié ou le taux de compensation salariale en cas d'absence pour maladie, que nous n'avons pas dans nos données, en sont deux exemples.

Pour évaluer si l'absence de contrôle de ces inobservables risque de beaucoup peser sur les résultats, nous avons estimé parallèlement un modèle de « sélection sur inobservables », appelé ainsi parce qu'il prend explicitement en compte l'existence possible de caractéristiques inobservables corrélées à la fois au type d'horaires et à l'absence pour maladie. Formellement, il conduit au système de 2 équations :

$$\begin{cases} hor\_irreg^* = x_1\beta_1 + u_1 \\ absence^* = x_2\beta_2 + \gamma \cdot hor\_irreg + u_2 \end{cases} \quad (13)$$

où  $absence^*$  et  $hor\_irreg^*$  sont des variables latentes, observées par les variables binaires repérant respectivement l'absence pour maladie (lors d'une des deux semaines de référence de l'enquête) et l'occupation d'un poste de travail en horaires irréguliers. Si des caractéristiques inobservées déterminent à la fois l'absence et l'irrégularité des horaires, alors les deux résidus  $u_1$  et  $u_2$  sont corrélés.

La difficulté avec ce type de modèle est de savoir si, pour en identifier les paramètres et notamment le paramètre  $\gamma$  mesurant l'impact de l'irrégularité des horaires sur l'absence pour maladie, il est nécessaire de poser une condition d'exclusion sur les variables. Cette condition consiste à exclure de  $x_2$  au moins une variable de  $x_1$ . Il n'y a pas de consensus chez les économètres sur cette question. Certains affirment que

dans le cas d'un probit (*i.e.* lorsque les résidus  $u_1$  et  $u_2$  sont supposés suivre une loi normale bivariée), elle n'est pas nécessaire à condition d'avoir suffisamment de variabilité dans les données (voir par exemple Wilde, 2000). D'autres soutiennent qu'en l'absence de restriction, les paramètres sont « faiblement » identifiés, c'est-à-dire ne le sont que sur les données, et que la méthode repose alors uniquement sur la bonne spécification de la loi des résidus.

Si on dispose d'une condition d'exclusion convaincante, c'est-à-dire si on est capable de trouver une variable qui agit sur le fait de travailler en horaires irréguliers mais n'a pas d'impact direct sur l'absence pour maladie, alors il faut retenir cette condition d'exclusion. Sinon, on a le choix entre deux options : poser tout de même une condition d'exclusion, même si elle est « à moitié » convaincante ; ou prendre les mêmes variables dans les deux équations, et compter sur la loi de distribution des résidus pour identifier les paramètres<sup>16</sup>.

Nous avons pris ici le parti de ne pas exclure de variable, faute de trouver condition d'exclusion valable. Notre modèle est donc « faiblement » identifié. Pour cette raison, il a ici le statut - annexe - d'un test de validité des estimations obtenues par la méthode d'appariement.

Pour obtenir, avec ce modèle, l'équivalent de  $\Delta$ , on part de la quantité :

$$\Pr(absence = 1 | hor\_irreg = 1) - \Pr(absence = 1 | hor\_irreg = 0) .$$

En supposant que le couple  $(u_1, u_2)$  suit une loi normale bivariée centrée réduite de coefficient de corrélation  $\rho$ , cette quantité s'écrit :

$$\frac{\Phi_2(x\beta_1, x\beta_2 + \gamma, \rho)}{\Phi(x\beta_1)} - \frac{\Phi_2(-x\beta_1, x\beta_2, -\rho)}{\Phi(-x\beta_1)} , \quad (14)$$

où  $\Phi$  (*resp.*  $\Phi_2$ ) est la fonction de répartition de la loi normale centrée réduite univariée (*resp.* bivariée). Une fois les paramètres de (13) estimés, on peut calculer (14) pour chaque salarié en horaires irréguliers. En prenant la moyenne empirique sur le sous-échantillon des ouvriers en horaires irréguliers, on obtient l'équivalent de  $\Delta$ .

---

<sup>16</sup> Le problème fondamental est que l'identification repose toujours sur une hypothèse non testable : dans le cas des méthodes d'appariement, sur l'indépendance des congés maladie au choix de travailler en horaires irréguliers, conditionnellement aux observables ; dans le cas des modèles de sélection sur inobservables, sur l'existence d'une variable d'exclusion (formellement identique à un instrument) ou sur l'adéquation des formes fonctionnelles retenues aux données.



## V - Résultats

### V.1 Résultats sur l'ensemble des ouvriers en horaires irréguliers (tous âges confondus)

La table 2 donne les résultats du logit expliquant l'irrégularité des horaires par les variables de la table 1<sup>17</sup>.

**Table 2 - Probabilité de travailler en horaires irréguliers (modèle logit)**

	<i>Paramètre estimé</i>	
<i>Constante</i>	-0.455	(0.391)
<i>Age</i>	-0.012***	(0.003)
<i>Age de fin d'études</i>	0.032***	(0.010)
<i>Possession d'un CAP-BEP</i>	-0.040	(0.046)
<i>Contrat à durée indéterminée</i>	0.081	(0.091)
<i>Ancienneté dans l'entreprise</i>		
moins d'un an	0.172*	(0.089)
de 1 an à moins de 5 ans	0.129**	(0.065)
de 5 ans à moins de 10 ans	0.119	(0.072)
10 ans ou plus	ref	-
<i>Catégorie socioprofessionnelle</i>		
ouvriers qualifiés de type industriel	-0.258	(0.237)
ouvriers qualifiés de type artisanal	-0.889***	(0.235)
chauffeurs	0.077	(0.245)
ouvriers qualifiés de la manutention ...	-0.236	(0.244)
ouvriers non qualifiés de type industriel	-0.367	(0.239)
ouvriers non qualifiés de type artisanal	-0.980***	(0.253)
ouvriers agricoles et assimilés	ref	-
<i>Fonction exercée</i>		
production, chantier, exploitation	0.099	(0.095)
installation, réparation, maintenance	-0.248**	(0.108)
nettoyage, logistique, ...	-0.315***	(0.090)
autres fonctions	ref	-
<i>Secteur d'activité</i>		
agriculture, sylviculture, pêche	-0.848***	(0.254)
industries agricoles et alimentaires	-0.084	(0.158)
industries des biens de consommation	-0.196	(0.169)
industrie automobile	-0.189	(0.172)
industries des biens d'équipement	-0.799**	(0.161)
industries des biens intermédiaires	0.034	(0.149)
construction	-1.347***	(0.146)
commerce et réparations	-0.626***	(0.145)
transports	0.389*	(0.152)
services aux entreprises	-0.166	(0.162)
services aux particuliers	ref	-
<i>Taille de l'entreprise</i>		
moins de 10 salariés	-0.607***	(0.149)
de 10 à 49 salariés	-0.474***	(0.144)
de 50 à 199 salariés	0.091	(0.145)
200 salariés ou plus	0.831***	(0.146)
inconnue	ref	-
<i>Accord de modulation ou d'annualisation du temps de travail (%)</i>	0.564***	(0.053)

Ecarts-type entre parenthèses. Seuils de significativité : \* : 10 % ; \*\* : 5 % ; \*\*\* : 1 %

Champ : ouvriers du secteur privé, âgés de 18 à 59 ans

Source : enquêtes Emploi en continu 1<sup>er</sup> trim. 2002 - 4<sup>ème</sup> trim. 2004.

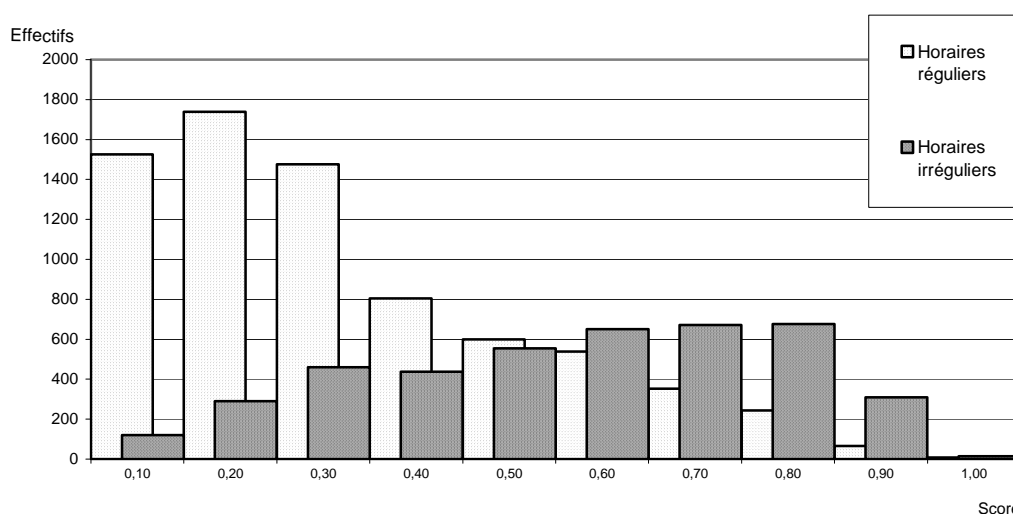
<sup>17</sup> L'âge a été introduit comme variable continue. Les indicatrices de région n'ont pas été reportées.

Le coefficient d'Estrella<sup>18</sup> vaut 23.3 %. Si on omet les indicatrices de secteur, il vaut 19.7 %. Si, de plus, on retire les indicatrices de taille d'entreprise, alors il baisse encore sensiblement et s'établit à 13,3 %. Ceci confirme le rôle très discriminant des caractéristiques d'entreprise dans l'irrégularité des horaires.

Avec la fonction de score estimée à partir du logit, l'estimation de  $\Delta$  peut se faire sur (presque) tout l'échantillon : seuls 2 ouvriers en horaires irréguliers, qui ont un score supérieur au score maximum des ouvriers en horaires réguliers, ont été éliminés. Le problème est que la fonction de score déduite du logit n'est pas « balancée ». Il n'y a pas d'équirépartition des salariés en horaires irréguliers et en horaires réguliers, conditionnellement au score de propension, dans un secteur d'activité (industries des biens intermédiaires) et dans 2 fonctions exercées (production ... et installation ...). On a donc enrichi le logit en croisant la variable « industries des biens intermédiaires » et les deux indicatrices de fonction concernées, avec les quatre indicatrices de taille d'entreprise, et on a en outre croisé l'indicatrice représentant la fonction *installation* ... avec des variables de secteur<sup>19</sup>.

La figure 2 représente la distribution de la fonction de score selon la régularité des horaires.

**Figure 2 - Distribution du score pour les salariés en horaires réguliers et les salariés en horaires irréguliers**



Lecture : 1 739 ouvriers en horaires réguliers ont un score compris entre 0,10 et 0,20.

Champ : ouvriers du secteur privé, âgés de 18 à 59 ans

Source : enquêtes Emploi en continu 1<sup>er</sup> trim. 2002 - 4<sup>ème</sup> trim. 2004.

On se trouve ici dans le cas très favorable où les deux sous-populations des salariés en horaires irréguliers et des salariés en horaires réguliers sont clairement distinctes (les modes des deux distributions sont espacés) et ont un support commun très étendu (il couvre quasiment tout l'échantillon). Cela signifie qu'il est toujours possible

<sup>18</sup> Estrella (1998) a proposé un indicateur d'ajustement pour le logit dichotomique, qui soit calibré comme le coefficient de détermination  $R^2$  du modèle linéaire standard : la valeur 0 correspond à une absence totale d'ajustement et la valeur 1 à un ajustement parfait.

<sup>19</sup> Agrégé au niveau de la NAF 4 : agriculture, construction, secteur industriel et secteur tertiaire.

pour un salarié en horaires irréguliers de trouver un salarié en horaires réguliers avec une valeur très proche du score de propension

Avec cette fonction de score, l'impact du fait de travailler en horaires irréguliers sur la probabilité de s'arrêter pour maladie s'élève à 1,23 points, avec un écart-type estimé par bootstrap (50 réplifications) de 0,55. La différence est cette fois significative (avec un seuil inférieur à 5 %). Ainsi, quasiment 20 %<sup>20</sup> de l'absence pour maladie chez les ouvriers en horaires irréguliers s'expliquerait par le fait même de travailler avec ce type d'horaires.

Pour conforter ce résultat, nous avons estimé le modèle probit bivarié<sup>21</sup> (13) sans exclusion de variable et en avons déduit l'équivalent de  $\Delta$  à partir des expressions (14). On obtient 1,28, valeur très proche de celle estimée par la méthode d'appariement, ce qui est plutôt rassurant.

## V.2 Résultats par âge

Nous avons repris toutes les estimations précédentes en les effectuant successivement sur 4 sous-populations de salariés distingués par leur âge. La table 3 récapitule les résultats. La colonne (1) reprend les taux d'absence constatés chez les salariés en horaires irréguliers. La colonne (2) donne la simple différence entre les taux d'absence calculés chez les ouvriers en horaires irréguliers d'une part, et chez les ouvriers en horaires réguliers d'autre part. La colonne (3) donne les résultats de l'estimation selon la méthode d'appariement sur le score de propension. Enfin, la colonne (4) donne les résultats de l'estimation par le probit bivarié (13).

**Table 3 - Influence des horaires irréguliers sur la probabilité de s'arrêter pour maladie**

Age	Nombre d'observations	Probabilité d'être en arrêt maladie (salariés en horaires irréguliers) (1)	Impact du travail en horaires irréguliers sur la probabilité de s'arrêter pour maladie		
			simple différence (2)	estimation par appariement (3)	estimation par probit bivarié (4)
Ensemble	11 538	6.21 (0.37)	0.65 (0.45)	1.23** (0.55)	1.28** (0.63)
18-29 ans	2 949	3.56 (0.57)	-0.39 (0.73)	-1.03 (1.34)	-1.05 (1.24)
30-39 ans	3 526	6.29 (0.67)	1.32 (0.79)	1.72 (1.06)	2.26* (1.18)
40-49 ans	3 049	7.36 (0.79)	1.69 (0.92)	2.80** (0.98)	3.17* (1.70)
50-59 ans	2 014	8.38 (1.07)	-0.24 (1.32)	2.74** (1.29)	1.94 (2.21)

Ecarts-type entre parenthèses. Seuils de significativité : \* : 10 % ; \*\* : 5 %.

Champ : ouvriers du secteur privé, âgés de 18 à 59 ans

Source : enquêtes Emploi en continu 1<sup>er</sup> trim. 2002 - 4<sup>ème</sup> trim. 2004.

<sup>20</sup> 1,23 / 6,21 = 0,198 (6,21 % est, rappelons-le, la proportion d'ouvriers en horaires irréguliers s'étant arrêtés pour maladie. A titre de comparaison, la CNAM estimait en 2003 à 6% les arrêts maladie injustifiés sur le plan médical (<http://www.ameli.fr/pdf/910.pdf>).

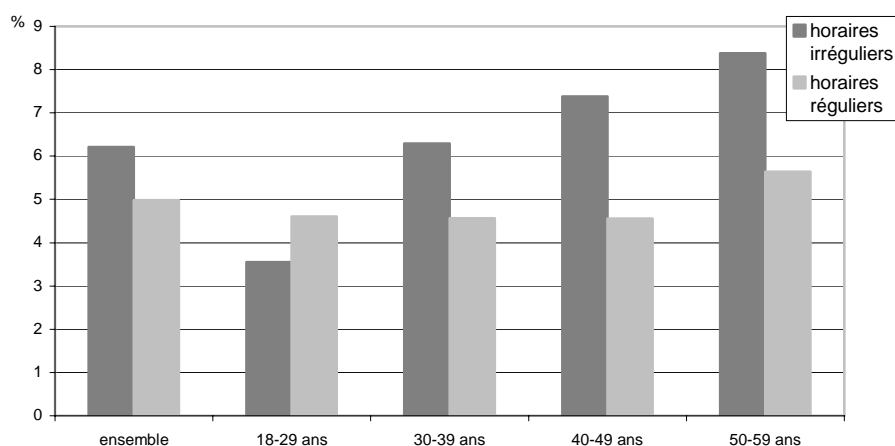
<sup>21</sup> L'estimation a été réalisée avec la commande *biprobit* de Stata.



La figure 3 ci-dessous est le pendant de la figure 1, où les probabilités d'absence observées chez les ouvriers en horaires réguliers de la figure 1 sont remplacées par les valeurs contrefactuelles déduites des colonnes (1) et (3) de la table 3.

La différence avec la figure 1 est frappante. D'abord, les écarts entre horaires irréguliers et horaires réguliers sont nettement plus marqués. Le cas de 50-59 ans est à cet égard particulièrement intéressant. La simple comparaison des taux d'absence chez les ouvriers en horaires irréguliers et chez les ouvriers en horaires réguliers montre a priori que les premiers sont un peu moins souvent malades que les seconds, même si la différence n'est pas significative (figure 1, construite avec la colonne (2) de la table 3). Mais si on tient compte de l'hétérogénéité observée, alors c'est l'inverse que l'on constate, avec cette fois-ci un écart significatif. Cette conclusion est confortée par les résultats du probit bivarié (colonne (4) de la table 3), certes très imprécis à cause de la faiblesse de l'identification du modèle (voir section 4.2 *supra*).

**Figure 3 - Taux d'arrêts maladie observés en horaires irréguliers et simulés en horaires réguliers**



Lecture : 6,3 % des ouvriers en horaires irréguliers se sont arrêtés pour maladie ; s'ils travaillaient en horaires réguliers, leur taux d'absence serait de 4,6 %.

Champ : ouvriers du secteur privé, âgés de 18 à 59 ans

Source : enquêtes Emploi en continu 1<sup>er</sup> trim. 2002 - 4<sup>ème</sup> trim. 2004.

Deuxième enseignement, il y a une réelle hétérogénéité de l'impact de l'irrégularité des horaires sur les comportements d'absence. Notamment, il est négatif (quoique non significatif) pour les 18-29 ans : chez les plus jeunes, l'effet « salaire » du cadre théorique l'emporterait (légèrement).

Dernier point, on ne note pas de variation, selon l'âge, du taux d'arrêt maladie qui prévaudrait en l'absence d'irrégularité des horaires, sauf chez les plus âgés qui s'arrêteraient un peu plus souvent.

## VI - Quelques remarques conclusives

Pour étudier le rôle des conditions de travail sur les absences pour maladie, nous avons proposé un modèle théorique fondé sur l'hypothèse que les conditions de travail ont un effet indirect sur les comportements d'absence qui transite par l'état de santé de l'individu. Dans ce cadre, l'impact de mauvaises conditions de travail sur l'absence pour maladie est ambigu, car il résulte de deux effets antagonistes : un effet « salaire » désincitatif à l'arrêt, dû au fait que les travaux pénibles sont mieux rémunérés, toutes choses égales d'ailleurs ; et un effet « santé » incitatif à l'arrêt, consécutif à la dépréciation plus rapide du capital santé de l'individu exposé à des conditions de travail pénibles.

Ce faisant, nous nous sommes écartés des modèles plus traditionnels qui traitent l'absence comme un phénomène d'aléa moral : profitant du manque d'information de l'employeur sur l'état de santé réel de son salarié, celui-ci utiliserait en partie la possibilité de prendre un arrêt maladie pour diminuer son effort au travail.

Nous n'ignorons pas l'existence de tels phénomènes, ni le fait que des conditions de travail moins favorables ont un impact direct sur la désutilité de ce dernier. Toutefois, nous considérons qu'il est difficile, voire impossible, d'identifier séparément dans les comportements d'absence ce qui relève de « purs » effets de santé et de « purs » effets d'aléa moral ou de désutilité du travail. Nous avons partiellement cherché à contrôler les effets d'aléa moral en utilisant des méthodes d'appariement : s'il se trouve que les déterminants - largement inobservables - des comportements d'aléa moral sont distribués également chez les ouvriers en horaires irréguliers et les ouvriers en horaires réguliers, alors ils n'invalident pas l'hypothèse d'indépendance conditionnelle (7) à la base de la méthode d'estimation. Certes, nous ne disposons pas dans nos données d'un des déterminants majeurs des comportements d'aléa moral, qui est le taux de compensation salariale des absences pour maladie. Mais la richesse des informations tant sur l'entreprise que sur les individus que nous avons utilisées, et, surtout, la comparaison des résultats avec un modèle qui prend en compte les caractéristiques inobservables qui risquent de peser sur les comportements donnent du crédit aux conclusions de l'étude.

Cela étant, l'étude a plusieurs limites. D'abord, nous avons été conduits, compte tenu des sources disponibles, à retenir un aspect particulier des conditions de travail, l'irrégularité des horaires. Nous avons écarté d'autres dimensions qui ont a priori un impact plus immédiat sur l'état de santé, celles qui sont liées aux pénibilités physiques et psychiques, ou aux risques professionnels. Nous ignorons en particulier si les postes de travail soumis à l'irrégularité des horaires sont aussi exposés à ces autres types de nuisances.

Plus fondamentalement, nous avons supposé dans le cadre théorique que l'indice  $p$  de conditions de travail était exogène (voir la section 2). Transposée au plan empirique, cette hypothèse signifie qu'il n'y a pas de phénomènes de sélection ou d'auto-sélection des individus dans les postes de travail reposant sur des déterminants individuels inobservés par l'économètre, et en premier lieu l'état de santé du salarié. Or ceci va à l'encontre d'un constat maintes fois rapporté dans la littérature épidémiologique, connu sous l'appellation « effet du travailleur sain » (*healthy worker effect*). L'état de santé pourrait jouer à plusieurs niveaux. D'une part, les salariés candidats aux postes de travail pénibles seraient a priori en meilleure santé que les autres. D'autre part, les salariés les moins résistants en emploi dans ce type de postes les quitteraient davantage que les autres, pour éventuellement être réaffectés dans des emplois moins pénibles. L'idéal aurait été de connaître l'état de santé du salarié. A défaut, nous sommes conduits à supposer que les salariés en horaires irréguliers et ceux en horaires réguliers ont *en moyenne* le même état de santé, conditionnellement aux caractéristiques observées.

Autre limite de l'étude : elle ignore l'aspect « demande de travail », par exemple le fait que l'employeur a la possibilité de se séparer de son salarié si celui-ci désorganise le processus de production à cause d'absences répétées<sup>22</sup>. Même s'ils ne peuvent être interprétés sans précaution comme des effets de demande, certains faits observés dans nos données vont dans ce sens. En effet, 11 % des salariés ayant déclaré s'être arrêtés pour maladie lors de la première vague de l'enquête n'étaient plus en emploi lors de la deuxième vague (trois mois après). Le pourcentage correspondant pour les ouvriers ne s'étant pas arrêtés est de 3,4 %.

Ceci conduit à une quatrième limite de l'étude. Elle a fondamentalement adopté un cadre statique, et ne capte pas les effets à moyen ou long terme des mauvaises conditions de travail sur la santé du salarié, qui conditionnent sa productivité et sa propension à rester en emploi.

Néanmoins, les premiers résultats obtenus nous semblent prometteurs et encouragent à poursuivre le travail. Le problème réside dans la disponibilité de sources de données adaptées, dont la collecte, on l'a vu, est très exigeante.

---

<sup>22</sup> Ce type de phénomène a été mis en évidence par Hesselius (2003) dans le cas de la Suède.

## Références

- d'Autume A., J.-P. Betbèze et J.-O. Hairault (2005), *Les seniors et l'emploi en France*, Rapport du Conseil d'Analyse Economique, 58, 241 p.
- Allen D. (1981), "An empirical model of worker attendance", *Review of Economics and Statistics*, 71, pp. 1-17.
- Andrén D. (2005), "Never on a Sunday : Economic Incentives and Sick Leave in Sweden", *Applied Economics*, 37, pp. 327-338.
- Askenazy P. et E. Caroli (2003), "New Organizational Practices and Well-Being at Work: Evidence for France in 1998", Research Unit Working Papers 0311, Laboratoire d'Economie Appliquée, INRA, 37 p.
- Barmby T.A., J.G. Sessions and J.G. Treble (1994), "Absenteeism, Efficiency Wages and Shirking", *Scandinavian Journal of Economics*, 96, pp. 561-566.
- Becker S. and A. Ichino (2002), "Estimation of average treatment effects based on propensity scores", *The Stata Journal*, 2(4), pp. 358-377.
- Bolduc D., B. Fortin, F. Labrecque and P. Lanoie (2002), "Workers' Compensation, Moral Hazard and the Composition of Workplace Injuries", *Journal of Human Resources*, 37, pp. 623-652.
- Bonhomme S. and G. Jolivet (2005), "The Pervasive Absence of Compensating Differentials", CREST Working Paper n° 2005-28, 54 p.
- Brown C. (1980), "Equalizing Differences in the Labor Market", *The Quarterly Journal of Economics*, vol 94, pp. 113-134.
- Brown S. and J.G. Sessions (1996), "The Economics of Absence: Theory and Evidence", *Journal of Economic Surveys*, 10, pp. 23-53.
- Caliendo M. and S. Kopeinig (2006), "Some Practical Guidance for the Implementation of Propensity Score Matching", *IZA Discussion Paper*, n° 1588, 29 p.
- Case A. and A. Deaton (2005) : "Broken Down by Work and Sex: How our Health Declines", in D. Wise (ed.), *Analyses in the Economics of Aging*, University of Chicago Press.
- Commission Européenne (2002), "Adapting to change in work and society: a new Community strategy on health and safety at work 2002-2006", COM(2002) 118 final, [http://europe.osha.eu.int/systems/strategies/future/com2002\\_en.pdf](http://europe.osha.eu.int/systems/strategies/future/com2002_en.pdf).
- Costa G. (1996), "The impact of shift and night work on health", *Applied Ergonomics*, 27(1), pp 9-16.
- Couffinhal A., N. Grandfils, M. Grignon et Th. Rochereau (2004), « La complémentaire maladie d'entreprise », *Question d'Economie de la Santé*, n°83.
- Dehejia R.H. (2005), "Practical propensity score matching: a reply to Smith and Todd", *Journal of Econometrics*, Vol. 125, Issue 1-2, pp. 355-364.
- Dunn L.F. and S.A. Youngblood (1986), "Absenteeism as a mechanism for approaching an optimal labour market equilibrium: an empirical study", *Review of Economics and Statistics*, 68, pp. 668-674.

Engelland A. and R. Riphahn (2005), "Temporary Contracts and Employee Effort", *Labour Economics*, 12(3), pp. 281-299.

Estrella A. (1998), "A New Measure of Fit for Equations with Dichotomous Dependent Variables", *Journal of Business and Economic Statistics*, 16(2), pp. 198-205.

Frick B. and M.A. Malo (2005), "Labour Market Institutions and Individual Absenteeism in the European Union", mimeo, Faculty of Management and Economics, Witten/Herdecke University, 38 p.

Gollac M. (1997), "Des chiffres insensés ? Pourquoi et comment on donne un sens aux données", *Revue française de sociologie*, n° XXXVIII-1, p.1-36.

Grignon M. and T. Renaud (2004), "Sickness and injury leave in France : moral hazard or strain ?", communication au 1er congrès franco-britannique en économie de la santé, Paris, 14-16 janv 2004.

Grossman M. (1972), "On the concept of health capital and the demand for health", *Journal of Political Economy*, vol. 80, n° 2, pp 223-255.

Heckman J.J., H. Ichimura and P. Todd (1998), « Matching as an economic evaluation estimator », *Review of Economic Studies*, 65(2), pp. 261-294.

Hesselius P. (2003), « Does Sick Absence Increase the Risk of Unemployment? », Working Paper Series, n° 2003-15, Department of Economics, Uppsala University, 32 p. (à paraître dans *The Journal of Socio-Economics*).

Lanfranchi J., H. Ohlsson and A. Skalli (2002), "Compensating wage differentials and shift work preferences", *Economics Letters*, vol 74, pp. 393-398.

Lasfargues G. (en coll. avec Molinié A.-F. et S. Volkoff) (2005), "Départs en retraite et travaux pénibles : l'usage des connaissances scientifiques sur le travail et ses risques à long terme pour la santé", *Rapport de recherche n° 19*, Centre d'Etudes de l'Emploi, 38 p.

Leuven E. and B. Sianesi (2003), "PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing", the Institute for Fiscal Studies.

Meyer B.D., W.K. Viscusi and D.L. Durbin (1995), "Workers' Compensation and Injury Duration: Evidence from a Natural Experiment", *American Economic Review*, 85(3), pp. 322-340.

Leigh J.P. (1991), "Employee and job attributes as predictors of absenteeism in a national sample of workers: the importance of health and dangerous working conditions", *Social Science and Medicine*, 33(2), pp. 127-137.

Midy F. (2005), "Les indemnités journalières versées au titre des arrêts maladie par le régime général : état des lieux et déterminants", *Revue médicale de l'assurance maladie*, Vol 3, pp 237-246.

Molinié A.-F. (2003), "Interroger les salariés sur leur passé professionnel : le sens des discordances", *Revue d'Épidémiologie et de Santé Publique*, 51, pp. 589-605.

Ose S. (2005), "Working conditions, compensation and absenteeism", *Journal of Health Economics*, 24(1), pp. 161-188.

Rosenbaum P.R. and D.R. Rubin (1983), « The Central Role of the Propensity Score in Observational Studies for Causal Effects », *Biometrika*, 72, pp. 41-55.

Rosen S. (1974), "Hedonic Prices and Implicit Markets: Product Differentiation in Pure Competition", *Journal of Political Economy*, Vol. 82, pp. 34-55.

Sickles R.C. and P. Taubman (1986), « An Analysis of the Health and Retirement Status of the Elderly », *Econometrica*, Vol. 54, pp. 1339-1356

Shapiro C. and J. Stiglitz (1984). "Equilibrium Unemployment as a Worker Discipline Device", *American Economic Review*, Vol. 74, pp. 433-444.

Smith J.A. and P. Todd (2005). "Does matching overcome LaLonde's critique of nonexperimental estimators?", *Journal of Econometrics*, Vol. 125, Issue 1-2, pp. 305-353.

Wagstaff A. (1986), "The Demand for Health: Some New Empirical Evidence", *Journal of Health Economics*, 5, pp. 195-233.

Wilde J. (2000), "Identification of multiple equation probit models with endogenous dummy regressors", *Economics Letters*, 69, pp. 309-312.

## Annexe : les macros SAS

Le logiciel Stata offre une riche bibliothèque de commandes permettant de mettre en œuvre ces méthodes d'appariement sur le score de propension (voir Becker et Ichino, 2002 ou Leuven et Sianesi, 2003). Si on ne dispose pas du logiciel Stata, on peut tout de même appliquer ces méthodes avec Sas grâce aux deux macros ci-dessous, qui ont été écrites pour l'occasion.

La macro `score` calcule le score, trace et conserve les deux courbes de densité du score pour les non traités et pour les traités avant sélection des individus sur le support commun, et donne la part des traités et des non traités qui sont exclus du support commun. Les paramètres de la macro sont les suivants :

- `tab_ent` : table en entrée contenant toutes les observations
- `y` : variable de résultat sur laquelle on estime l'effet du traitement
- `d` : variable binaire de traitement ; elle doit valoir 0 pour les non traités et 1 pour les traités
- `x` : liste des variables explicatives du score
- `score` : variable (nommée `score` par défaut) contenant le score
- `print_logit` : option d'impression (pas d'impression par défaut) des résultats du logit
- `print_graph` : option d'impression (pas d'impression par défaut) des deux courbes de densité du score
- `tab_graph` : table (nommée `t_graph` par défaut) contenant les données sur les deux courbes de densité ; cette table peut être transformée en table Excel, puis en graphique Excel ; pour la récupérer, le paramètre `print_graph` doit être à 1
- `taux_exclus` : taux (en centiles) d'observations exclues des groupes des traités et des non traités lors de la détermination du support commun ; lorsque le paramètre est à 0, aucune observation n'est exclue, et les minima et maxima des deux distributions sont directement comparés pour la détermination du support commun ; lorsque le paramètre vaut 1 (*resp.* 5), 1% (*resp.* 5 %) des observations sont éliminées aux deux extrêmes des deux distributions, et ce sont ensuite les minima et maxima de ces distributions tronquées aux deux bouts qui sont comparés pour la détermination du support commun ; seules trois valeurs du paramètre `taux_exclus` sont possibles : 0 (par défaut), 1 et 5.
- `exclus_inf` : vaut 0 par défaut, auquel cas aucun non traité n'est exclu du support commun
- `print_exclus` : option d'impression (pas d'impression par défaut) du pourcentage des individus exclus du support commun
- `tab_sort` : table en sortie contenant toutes les observations du support commun et les variables `&y`, `&d`, `&x`, `&score` (voir *supra*) et `&var_sup` (voir *infra*)
- `var_sup` : liste des variables autres que `&y`, `&d`, `&x`, `&score` à conserver dans la table de sortie des observations du support commun

```
%macro score(tab_ent= ,
              y= , d= , x= ,
              score=,
              print_logit=0,
              print_graph=1,tab_graph=t_graph,
              taux_exclus=0,exclus_inf=0,print_exclus=0,
              tab_sort=,
              var_sup=);
  /***** calcul du score */
proc logistic data=&tab_ent descending
              %if &print_logit=0 %then noprint;;
  model &d = &x ;
  output out=tab_score p=&score;
run;
  /***** visualisation des scores*/
%if &print_graph=1 %then %do;
proc kde data=tab_score(where=(&d=0)) method=os bwm=1
  out=_s0(keep=&score density)
  levels=1 percentiles=1; *options pour minimiser les impressions;
var &score;
```

```

run;
proc kde data=tab_score(where=(&d=1)) method=os bwm=1
    out=_s1(keep=&score density)
    levels=1 percentiles=1; *options pour minimiser les impressions;
    var &score;
run;
data &tab_graph;
    merge _s0(rename=(density=d0)) _s1(rename=(density=d1));
    by &score;
run;
proc gplot data=&tab_graph;
    symbol i=join;
    plot (d0 d1)*&score/overlay;
run;
quit;
%end;
/***** exclus du support commun */
proc sort data=tab_score;
    by &d;
run;
proc univariate data=tab_score noprint;
    by &d;
    var &score;
    %if &taux_exclus=0 %then %do;
        output out=tab_min_max(keep=&d p_min p_max)
            min=p_min max=p_max;
    %end;
    %if &taux_exclus=1 %then %do;
        output out=tab_min_max(keep=&d p_min p_max)
            p1=p_min p99=p_max;
    %end;
    %if &taux_exclus=5 %then %do;
        output out=tab_min_max(keep=&d p_min p_max)
            p5=p_min p95=p_max;
    %end;
run;
proc summary data=tab_min_max;
    var p_min p_max;
    output out=support(keep=supp_min supp_max)
        max(p_min)=supp_min
        min(p_max)=supp_max;
run;
data &tab_sort(keep=&y &d &x &score &var_sup)
    exclus(keep=&d exclus_inf exclus_sup);
    if _n_=1 then set support;
    set tab_score;
    %if &exclus_inf=0 %then %do;
        exclus_inf=0;exclus_sup=(supp_max<&score);output exclus;
        if &score<=supp_max then output &tab_sort;
    %end;
    %else %do;
        exclus_inf=(&score<supp_min);exclus_sup=(supp_max<&score);output exclus;
        if supp_min<=&score<=supp_max then output &tab_sort;
    %end;
run;
%if &print_exclus ne 0 %then %do;
proc summary data=exclus;
    class &d;
    var exclus_inf exclus_sup;
    output out=p_exclus(rename=(freq=effectifs)) sum= ;
run;
data p_exclus;
    set p_exclus;
    p_exclus_inf=exclus_inf/effectifs;
    p_exclus_sup=exclus_sup/effectifs;
proc print data=p_exclus noobs;
    var &d effectifs exclus_inf p_exclus_inf exclus_sup p_exclus_sup;
    label exclus_inf="exclus_inf (nbre)"
        p_exclus_inf="exclus_inf (%)"
        exclus_sup="exclus_sup (nbre)"
        p_exclus_sup="exclus_sup (%)";
run;
%end;
%mend;

```



La macro **kernel\_matching** calcule l'estimateur par appariement par la méthode du noyau et imprime les résultats du test de la *balancing property*. Les paramètres sont les suivants :

- **tab\_ent** : table en entrée, issue de la macro *score* (paramètre *tab\_sort*), contenant toutes les observations du support commun
- **y** : variable de résultat sur laquelle on estime l'effet du traitement
- **d** : variable binaire de traitement
- **x** : variables explicatives du score
- **x\_ref** : modalités de référence des éventuelles variables polytomiques du score ; ces modalités ne sont pas introduites dans le logit du score ; en revanche, elles sont utilisées pour le test de la *balancing property*
- **score** : variable de score, issue de la première étape de construction du score (*cf* macro *score*)
- **print\_estim** : option d'impression du résultat de l'estimation (pas d'impression par défaut) ; cette option est utile en cas de bootstrap
- **tab\_sort** : table de sortie, créée si *print=0*, contenant simplement la valeur de l'estimateur (variable *estim*) ; option à utiliser en cas de bootstrap
- **test\_balance** : option de test (pas de test par défaut) de la *balancing property*
- **print\_test** : option d'impression (pas d'impression par défaut) des résultats du test de la *balancing property*.

```
%macro kernel_matching(tab_ent= ,
                        y= , d= ,x= ,x_ref= ,
                        score=score,
                        print_estim=0,tab_sort=,
                        test_balance=0,print_test=0);

proc iml;
  /*** module de lecture des données */
  start lecture;
  use &tab_ent;
  read all var {&y &d &score &x} into don;
  /* traités */
  y_1=don[loc(don[,2]=1),1];
  score_1=don[loc(don[,2]=1),3];
  x_1=don[loc(don[,2]=1),4:ncol(don)];
  nobs_1=nrow(y_1);
  /* non traités */
  y_0=don[loc(don[,2]=0),1];
  score_0=don[loc(don[,2]=0),3];
  x_0=don[loc(don[,2]=0),4:ncol(don)];
  nobs_0=nrow(y_0);
  /* modalités de référence des variables polytomiques */
  %if &test_balance=1 %then %do;
    read all var {&d &x_ref} into don_ref;
    x_ref_1=don_ref[loc(don_ref[,1]=1),2:ncol(don_ref)];
    x_ref_0=don_ref[loc(don_ref[,1]=0),2:ncol(don_ref)];
  %end;
finish lecture;
  /*** module de construction du contrefactuel de chaque traité */
  start cf;
    /* calcul de la fenetre h du noyau */
    score_moy=sum(score_0)/nobs_0;
    score_centre=score_0-(j(nobs_0,1,1)*score_moy);
    variance_score=t(score_centre)*score_centre/nobs_0;
    sigma_score=sqrt(variance_score);
    h=sigma_score/(nobs_0##0.2);
    /* contrefactuel de i */
    cf=j(nobs_1,1,0);
    %if &test_balance=1 %then %do;
      x_cf=j(nobs_1,ncol(x_0||x_ref_0),0);
    %end;
    do i=1 to nobs_1;
      dif_score=(score_0-(j(nobs_0,1,1)*score_1[i]))/h;
      k=exp(-0.5#dif_score#dif_score); * poids;
      cf[i]=(k#y_0)[+]/k[+]; * contrefactuel;
      %if &test_balance=1 %then %do;
        x_cf[i]=(k#(x_0||x_ref_0))[+,]/k[+];
      %end;
    end;
  finish cf;
endmacro;
```

```

end;
finish cf;
/** estimation */
run lecture;
run cf;
estim=sum(y_1-cf)/nobs_1;
%if &print_estim=1 %then %do;
    print estim;
%end;
%else %do;
    create &tab_sort(rename=(coll=estim)) from estim;append from estim;
%end;
%if &test_balance=1 %then %do;
    x_1=x_1||x_ref_1;
    create x_1 from x_1[colname={&x &x_ref}];append from x_1;
    create x_cf from x_cf[colname={&x &x_ref}];append from x_cf;
%end;
quit;
run;
%if &test_balance=1 %then %do;
    data test_;
        set x_1(in=i1) x_cf(in=i2);
        if i1 then d=1;
        if i2 then d=0;
    run;
    ods listing close;
    proc ttest data=test_;
        class d;
        var &x &x_ref;
        ods output Statistics=tab_stat_ Ttests=tab_test_;
    run;
    data tab_stat_0(keep=variable mean
                    rename=(mean=moy_0));
        set tab_stat_;
        if class='          0';
    run;
    proc sort data=tab_stat_0;
        by variable;
    run;
    data tab_stat_1(keep=variable mean
                    rename=(mean=moy_1));
        set tab_stat_;
        if class='          1';
    run;
    proc sort data=tab_stat_1;
        by variable;
    run;
    data tab_test_m(keep=variable probt);
        set tab_test_;
        if method='Satterthwaite';
    run;
    proc sort data=tab_test_m;
        by variable;
    run;
    data tab_;
        merge tab_stat_0 tab_stat_1 tab_test_m;
        by variable;
    run;
    ods listing;
    ods output close;
    %if &print_test=1 %then %do;
        proc print data=tab_ noobs split='$';
            var variable moy_1 moy_0 probt;
            label moy_1="moyenne sur $ traités"
                  moy_0="moyenne sur $ non traités $ appariés"
                  probt="seuil de $ significativité";
        run;
    %end;
%end;
%mend;

```

